DOCUMENT RESUME

ED 022 286 40 EC 002 464

By-Walker, Hill M.; Mattson, Robert H.
IDENTIFICATION AND TREATMENT OF SOCIAL-EMOTIONAL PROBLEMS. INTERIM REPORT.
Oregon Univ., Eugene. Coll. of Education.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-6-1308
Pub Date May 67
Grant-OEG-4-6-061308-0571
Note-118p.
EDRS Price MF-$0.50 HC-$4.80
Descriptors-*BEHAVIOR, BEHAVIOR CHANGE, BEHAVIOR PATTERNS, BEHAVIOR RATING SCALES, CHECK LISTS, *EMOTIONALLY DISTURBED, *EXCEPTIONAL CHILD RESEARCH, *IDENTIFICATION, OBSERVATION, POSITIVE REINFORCEMENT, REINFORCEMENT, SCREENING TESTS, *TEACHING METHODS, TEST CONSTRUCTION, TEST RELIABILITY, TEST VALIDITY

Three assessment instruments for disturbed children were developed: a 50-item behavior checklist which functioned as a screening device; a 124-item behavior rating scale which provided frequency measures on indices of the teacher's reaction and response to exhibited behaviors; and a behavioral observation form which recorded task-oriented behavior in 10-second intervals for 10-minute periods. The checklist had a split-half reliability of .98 and discriminated between disturbed and non-disturbed children (p=.001). The rating scale reflected treatment differences which were known to exist (p=.01), and had an average inter-rater reliability of .935 for three judges on the behavior of six subjects. Agreement measures between independent observers using the observation form were .90 and above. A treatment model based upon learning theory was developed to modify the behavior of disturbed children in an educational setting. Various response-reinforcement contingencies and reinforcers were used with 11 disturbed boys in grades 4, 5, and 6 and produced measurable change by reducing deviant behavior and increasing time spent engaged in task-orientated behavior. It was not possible to determine which treatment variables produced a given amount of behavior change. The checklist, rating scale, and a classification form are appended. (Author/SN)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR-6-1308
P.A. 40

COVER PAGE

INTERIM REPORT

Grant No. OEG 4-6-061308-0571

Identification and Treatment of Social-Emotional Problems

May, 1967

U. S. Department of
Health, Education, and Welfare

Office of Education
Bureau of Research

ED022286

ED012964

TITLE PAGE

IDENTIFICATION AND TREATMENT OF SOCIAL-EMOTIONAL PROBLEMS

USOE Contract No. OEG 4-6-061308-0571

Hill M. Walker
Robert H. Mattson

May, 1967

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Oregon

Eugene, Oregon

NATURE OF THE PROBLEM

I. Introduction

During the past several years, there has been increasing attention directed toward the general problems of chronic malfunctioning and particularly toward behavior disturbance in children. Research has demonstrated that great numbers of children are non-functional or semi-functional in a behavioral sense. Many of these children are viewed as restricted in development. Others have been denied legitimate access to behavior patterns which are tolerated or preferred by the dominant culture and thus seek reinforcement by means of deviant behavioral functioning. Essentially, they have learned inappropriate or unacceptable behaviors.

Concomitant with the increasing attention to the problem of behavior disturbance has been the concern for developing therapeutic procedures for coping effectively with this problem. Until recently, mental health clinics and psychiatric facilities have assumed the primary responsibility for providing therapeutic treatment for these children. It has become apparent, however, that in order to meet the treatment needs of disturbed children, additional provisions, both educational and therapeutic, must be established for them. The public schools have the potential for fulfilling this role if effective techniques and strategies for modifying disturbed behavior can be implemented within the school setting. Such a procedure would increase the school system's holding power within disturbed populations of school children and would relieve some of the burden presently sustained by clinical treatment facilities.

1. Background of Information

The traditional conceptualization approaches to behavior disturbance among children have been described and classified by the Committee of Child Psychiatry of the American Psychiatric Association. (1966) The educationally relevant sections of this classification system can be described under three general categories: (A) Developmental deviations (B) Deviations in Social Development (C) Reactive Disorders.

Developmental deviations refer to behavior which is judged atypical for a particular age group in that it is not ordinarily expected for a given stage of development. For example, persistent babbling behavior may be appropriate for an eighteen month old child but quite inappropriate for a seven year old child. Thus, a child's behavior may be normal at one given age, yet judged inappropriate or deviant at another point in time due to a maturational lag in behavior.

Deviations in social development are reflected in the disparity which exists between that which is theoretically expected for a particular child and that which is empirically observed in that child. Within this schema, the child's social behavior is judged against his individual background, his abilities, and his personality configuation as opposed to age, sex, and grade norms. The size of the disparity between these variables and the number of such disparities represent magnitude and frequency measures of the child's behavior.

Reactive disorders refer to deviant behaviors and symptoms, exhibited by the child, which are judged to be primarily a reaction to an event, a set of events, or a situation. This symptomatic approach views behavior disturbance as the individual's response to pressing needs which are not being met by his environment or needs which the individual is unable to satisfy within his present social milieu.

All three of these conceptual approaches carry implications for identification and treatment. Each is concerned with isolating the disturbing behavior and establishing appropriate treatment procedures for effective remediation of the identified problem(s). Yet many identification and treatment strategies based upon such conceptual approaches have failed to cope effectively with behavior problems within public school populations. They have failed not primarily because of methodological or philosophical defects but because they have not established workable, correlational relationships between identification and treatment variables.

Although the need for early identification of behaviorally disturbed children is obvious, it is not practical to identify children for whom treatment does not exist nor to identify more children than existing treatment facilities can accommodate. Equally important in this regard is the need for developing identification criteria that are closely related to existing treatments to which children, so identified, can be referred. The categorization of behaviors into a classification system which will be prescriptive for treatment becomes crucial in the early stages of developing appraisal instruments for the identification of behaviorally disturbed children. Such categorization strengthens the relationship between identification and treatment variables and provides relevant information about which behaviors in a given item pool cluster together as related behaviors. Within this kind of framework, one can better argue that an individual with behavior cluster IA is most suited for existing treatment 2Y which has been designed for treatment of that specific sample of behavior.

2. Purpose

The purpose of the current research study is to construct and validate a multi-dimensional model which will be used for the identification, prediction, and assessment of the deviant behavior of disturbed children. The model is composed of three scales which represent increasingly refined levels of observation and assessment. The text of the report will focus upon the development and initial testing of these instruments and upon preliminary findings as revealed by appropriate data analyses.

3. Theoretical Rationale Underlying Scale Development Procedures

The behavior of disturbed children can be conceptualized as a behavioral domain which is composed of a number of interrelated variables. The actual nature of these variables will depend upon one's definition of disturbed behavior. Since disturbed behavior is a multiply-caused phenomenon, and not an isolated entity, the operations of measurement would have to sample an almost limitless number of relevant variables in order to adequately describe this behavioral domain.

A more feasible approach to this problem is to develop an operational definition of disturbed behavior, select component variables which are interrelated and which bear a direct, functional relationship to disturbed behavior, and to construct behavioral items that will objectively measure each of these component variables. However, this approach imposes two limitations upon the operations of measurement. First, it assumes that the totality of the particular domain can be described precisely and secondly that all items or contrived situations are equally effective measures of the component variables. (Ghiselli, 1964). It is very unlikely that the totality of a global concept such as disturbed behavior can be precisely described. Yet it is possible to specify the most relevant dimensions of disturbed behavior and to quantify these dimensions through the measurement process. It is also obvious that some items are more effective measures of preselected component variables, which comprise disturbed behavior, than are other items. Such items, then, are not equally representative sub-samples of the total behavioral domain. Thus, in the construct of domain sampling, one is often forced to shift from a random to a systematic selection of item measures. One systematically selects those items which, on subjective bases, are the most adequate measures of a given component variable. More objective information, relative to this relationship, is usually provided by item analysis techniques.

In the current project, disturbed behavior has been operationally defined as those overt, inappropriate behaviors which produce a reinforcing effect upon the environment. This definition suggests that maladaptive, disturbed behavior is learned in the same fashion as are adaptive, constructive modes of behavior. The relevant dimensions of disturbed behavior which are being investigated through quantification techniques in the current project are measures of: (1) The presence or absence of overtly, disturbing behavior (2) The frequency of overtly disturbing behavior (3) The environmental response to overtly, disturbing behavior (4) The teacher's reaction to overtly, disturbing behavior (5) The amount of task oriented behavior contained in the disturbed behavior pattern.

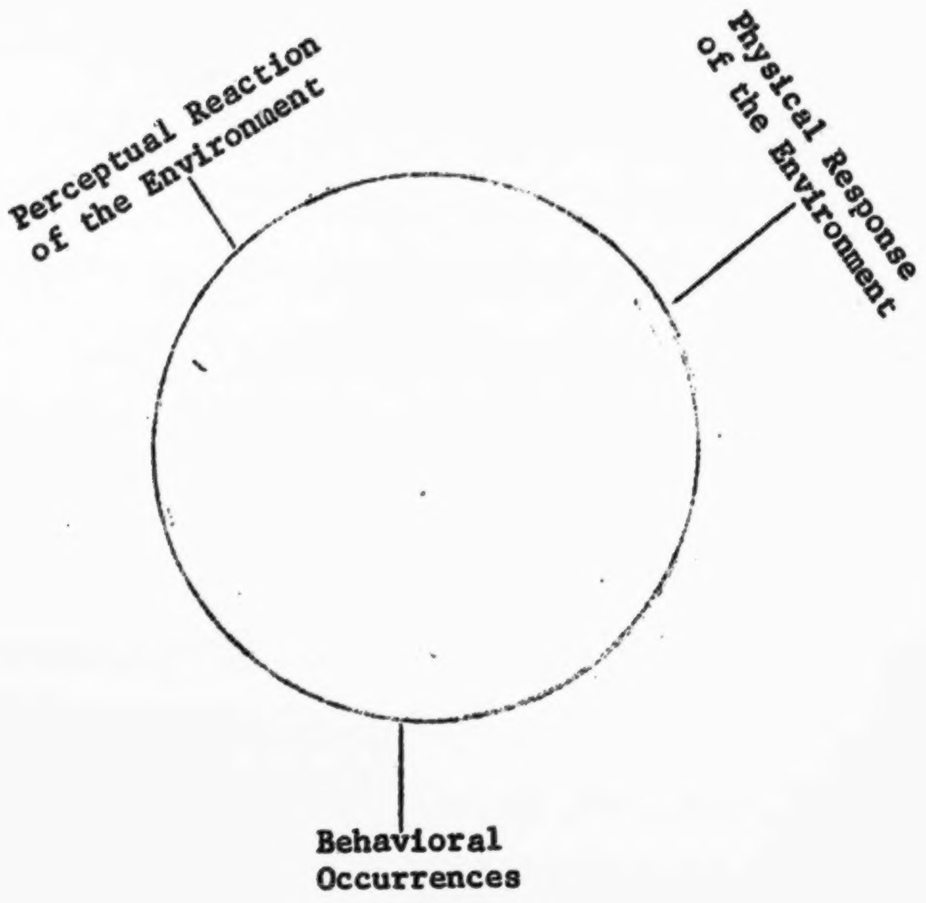
It was judged important in the current project to select items for measuring disturbed behavior that could be verified by observation. Inferences about given behavioral entities were held to a minimum in the measurement process because of the reliability attenuation which obtains when inferential judgments are elicited across observers. Therefore, an item pool of observable statements about behavior was collected for the purpose of building behavioral assessment instruments.

A simple measure of the presence or absence of a behavior or trait is judged by some test constructors as preferable to the more complex and refined scaling methods of rating scales. Thorndike and Hagen (1961).

Although this statement certainly applies to the measurement process in terms of inter-rater reliability, a simple measure of presence or absence reveals little information about the current status of a behavior. A measure of the frequency with which a behavior occurs is critical in assessing the status (stable or fluctuating) of a behavior as a function of a treatment process, a major environmental change, or other conditions which can be either specified or controlled.

Behavior is learned and regulated as a consequence of its effect upon the environment. It has been argued that deviant and non-deviant behaviors are acquired as a result of the same learning process. Patterson (1965). It seems probable that settings or environments differ in the manner in which they respond to operant behavior. For example, in one type of setting, the probability may be high that a given subject will learn appropriate behaviors. In another type of setting, the probability may be high that the same subject will learn inappropriate or deviant behaviors. Therefore, in assessing deviant behavior, it becomes necessary to evaluate the manner in which that behavior is reacted to by the environment. By its response to the behavior, the environment may operate to either reinforce or extinguish that behavior. Assessment procedures have been established in the current project to measure the way in which the school environment responds to deviant behavior in children.

For purposes of assessment in the current project, deviant behavior is seen as a product of the interaction between three sets of interrelated variables.


Perceptual Reaction of the Environment
Physical Response of the Environment
Behavioral Occurrences

The perceptual reaction of the environment is related to the occurrence of behavior and the environmental response to that behavior by way of the definition of disturbed behavior as not representing an isolated phenomenon. Behavior does not occur in a vacuum, and a teacher's perceptual reaction to an individual deviant behavior can affect the environment's response to that behavior. The teacher's reaction to the child's behavior, for example, may be a contributing factor in the eventual classification of that behavior as disordered or disturbed. Tolerance levels among teachers for different forms of abberant behavior is subject to considerable variation. The probability that a given child's disturbing behavior will eventually be classified as disordered is contingent upon his teacher's tolerance level for his particular type of disturbing behavior. Mattson, Mattos, Walker (1967). A measure of rater reaction has been built into the current study in order to collect data on the differential perceptual reactions of teachers to the disturbing behaviors exhibited by school children.

The variable of task-oriented behavior has been selected as the criterion measure of behavioral change as a function of the treatment process in this project. Task oriented behavior was selected for this purpose for several reasons. First, task oriented behavior, as defined by its measuring instrument, is incompatible with deviant or disturbed behavior. Thus, the presence of task oriented behavior indicates the absence of deviant behavior _during the interval in which task oriented behavior is being recorded_. Second, task oriented behavior is one of the most critically important variables operative within the school setting. The school is oriented toward fostering task oriented behavior, and it accordingly regards this role as one of its most important functions. The presenting problem in many educational referrals is non-attending or non-task oriented behavior, regardless of the numerous psychosocial correlates which are attributed to the referral. Since so many deviant classroom behaviors are contingent upon the presence or absence of task oriented behavior, the use of this variable as a criterion measure seemed most appropriate in this study.

The assessment instruments which have been developed in the current project represent a three stage process in which each succeeding stage provides for a more refined observation of disturbed behavior. The instruments consist of: (1) a fifty item behavior checklist which functions as an initial screening device, (2) a one hundred twenty-four item behavior rating scale which provides frequency measures on individual items and indices of the teacher's reaction and response to exhibited behaviors, (3) a behavioral observation form which records task oriented behavior in cumulative ten second intervals for ten minute observation periods. The next three sections of the interim report describe the development (and validation, where applicable) of these three instruments. Section I is devoted to the behavior checklist; Section II, to the behavior rating scale; and Section III, to the behavioral observation form. Section IV discusses procedures and results of the application of a behavior treatment model to six disturbed children in Treatment Phase II.

SCALE DEVELOPMENT PROCEDURES

Section I

Behavior Checklist

Reprinted from "Construction and Validation of a Behavior Checklist for the Identification of Children with Behavior Problems." By: Hill Montague Walker by permission of: Hill Montague Walker. Copyrighted: March, 1967. Copyrighted material includes Pages 11 through 53 and material included in Appendixes A and B.

REPRESENTATIVE REVIEW OF RELEVANT LITERATURE

As stated in the above heading, this will be a representative review of the literature. Literature will be considered here which bears upon teacher identification of children with behavior problems, score weighting methods, and related behavior checklists.

I. Teacher Identification of Children with Behavior Problems

The value and need for early identification of children with behavior problems seems to be generally accepted by educational and psychological personnel. However, since the publication of Wickman's 1928 Monograph comparing teachers' and clinicians' attitudes toward the behavior problems of children, the teacher's role in attempts at early identification has been viewed with some equivocation. Wickman found discrepancy rank order correlations of -.22 and -.11 respectively between the rankings of teachers and those of thirty mental health specialists on the relative seriousness of various problem behaviors of school children. Clinicians viewed withdrawal and other antisocial forms of behavior as more serious, in terms of pathology, than did teachers. On this same dimension, teachers were more concerned with behaviors disruptive of classroom order, discipline, and achievement. Wickman (1928).

Since the judgments of psychologists in this study were accepted as a criterion against which teacher judgment was compared, the variance between the judgment of these two groups raised serious questions about the competence of teachers in identifying disturbed children. It should be noted that Wickman's methodology has drawn considerable criticism which has cast some doubt upon the credibility of his findings. Watson (1933), for example, points out that teachers and clinicians were given differential instructions for the rating/ranking task in this study. Teachers were instructed to rank behaviors for present seriousness while clinicians were asked to rank them according to their importance or influence in handicapping a child's future adjustment.

Stouffer, (1952) reported a study in 1952 in which he used essentially the same research design as Wickman. This study showed a much closer agreement, positive rho of .61, between teachers and mental hygienists in their ranking of the relative seriousness of children's behavior problems. In addition, Stouffer found a rank order correlation of .87 between his and Wickman's mental hygienists Stouffer concluded that teachers' attitudes toward children's behavior problems had changed considerably since Wickman's study and had become more like those of psychologists.

Studies by Hunter (1957) and Ullmann (1952) were also reported in the fifties which showed greater congruence between teachers and mental health experts in their evaluations of childhood behavior problems than was the case at the time of Wickman's study. Schrupp and Gerde (1953), using the same research design as Wickman, found much more agreement between teachers and clinicians than was indicated in the late 1920's. However, the authors qualify this finding by pointing out that definite disagreements were still evident, and that the direction of the disagreements was similar to that found by Wickman. Schrupp and Gerde observe that, "Teachers, when compared with clinicians, still appeared to be less concerned about behavior traits associated with withdrawal and more concerned about those which appear to be transgressions against orderliness and perhaps morality." p.6

An opposite point of view is reflected in studies reported by Clark (1951) and Peck (1955). Peck's study revealed that teachers viewed undesirable personality traits as the most serious of behaviors; regressive traits were slightly less serious; and agressive behavior as rated least serious. Clark concludes from the results of his study that teachers are more disturbed by children's behaviors which annoy other children than by behaviors which affect teachers directly.

In the early sixties Sarason (1960) and his associates maintained that developing personality measures to identify children whose high anxiety level is interfering with a productive use of their potential is important because teachers do not perform this function to a satisfactory degree. These authors believe that teachers do not have the time or the training to act as psychological diagnosticians.

A different position is taken by Bower (1958) who used clinicians' judgments of emotionally disturbed children as a criterion against which he compared teachers' judgments of the same sample in terms of emotional disturbance. Bower found a very close relationship between teachers' and clinicians' judgments of disturbed behavior. Teachers identified eighty-seven percent of clinically identified children and identified a greater number of children as being overly withdrawn or timid than as overly aggressive or defiant. Evidence growing out of this study seems to refute the oft cited criticism that teachers tend to ignore withdrawn children whose behavior may not be as disruptive or disturbing as that of an acting out, aggressive child.

Beilin (1959) has summarized research since 1927 which relates to the validity of teachers' identification of children with social-emotional problems. His interpretation of research findings suggests that teachers have become more like clinicians in making judgments about children. Beilin believes teachers and clinicians will likely always differ in basic attitude. Teachers, because they are task oriented will probably focus more on problems disruptive of achievement than will clinicians. Clinicians, more concerned with adjustment, are more likely to identify withdrawn children who may be achieving satisfactorily.

Maes (1966) has reported a study which demonstrates that emotionally disturbed children in grades four, five, and six can be identified as effectively through the use of a teacher rating scale and a group intelligence test as through the use of these two sources of information in addition to arithmetic achievement, reading achievement, a modified sociometric technique (a class play), and a self-concept inventory. The predictive efficiency which Maes achieved with two variables (teacher rating and intelligence) equalled that which Bower achieved through the use of six variables. This procedure makes the identification process considerably more efficient and lends further support to Bower's finding that teacher judgment is an important variable in the identification of emotionally disturbed children.

Mathew Trippe (1961), in discussing the role of the teacher in identifying emotionally disturbed children, argues that competent teachers are the best judges of disturbed behavior in schools. He notes that requiring the judgments of teachers to be validated against the judgments of clinicians fails to recognize that the role of teaching is different from the role of treating. The failure to distinguish between these roles has resulted in some concern that teachers might indiscriminately label children as disturbed; however, he suggests that if a variety of school patterns were available, teachers' attention to disturbing children would result not in the treatment of an illness but in a better placement for the child.

Thus evidence exists that teachers are, at present, in much closer agreement with mental health specialists in their judgments of childhood/behavioral difficulties than was true thirty years ago. Although some questions are still raised about the validity of teacher judgment of childhood adjustment problems, there is a recognition that the classroom teacher's vantage point is an especially good one for such identification.

In this study, an attempt was made to combine judgments of clinicians and teachers about behavior problems in the construction of the behavior checklist. Operational statements abstracted from teachers' descriptions of problem children were rated by a five member behavioral science panel according to their influence or weight in handicapping a given child's behavioral adjustment. Ratings of the panel members were pooled and weights assigned to items on the basis of these ratings.

II. Score Weighting Methods

In the last twenty years there has been a fair amount of controversy over the utility of mathematically derived item weights. Research has consistently yielded high, positive correlations between mathematical weights, derived from the predictive significance of items, and arbitrarily assigned score weights, thus casting some doubt upon the value of computing weights mathematically in test construction. Levitt (1961), for example, states that, "Considerable, well-designed research has demonstrated that the correlation between scores yielded by an inventory with items weighted on an empirical, mathematical basis, and with arbitrary weighting, is around .90 or higher." p. 81. Levitt thus concludes that the methods are interchangeable and recommends using the simpler one.

Research evidence supportive of the above statement is reported by Guilford in Psychometric Methods. He cites empirical studies of weighting with tests by Guilford, Lovell, and Williams (1942), by Phillips (1954), and by Harper and Dunlap (1954). He notes that, "The first two found practically no improvement in reliability of achievement tests using weighted items. The third study substituted weights of -1, 0, and +1 in fourteen keys of the Strong Vocational Interest Blank for Women for the standard weights of -4 to +4. The new experimental scores correlated .95 to .99 with standard scores." pp. 14-21.

The research study by Guilford, Lovell, and Williams was designed to determine whether an examination with completely weighted scoring yields any more highly reliable and valid scores than with unweighted scoring and to determine whether length of examination has any bearing upon the effect of weighted versus unweighted scoring. Results of the study showed that weighted scoring yielded an average gain of .02 in reliability coefficients. In validity coefficients, the weighted scoring yielded a gain of .02 in the shortest test and less than this amount in the longest test--neither gain being statistically significant. Guilford (1942) concludes that, "The customary unweighted scoring which takes distinctly less time and effort gives about as reliable and valid results as differential weights afford." p.21.

Phillips (1943) reported a study in which he compared the unweighted or right-wrong scoring method with the weighted method on an intelligence test. No statistically significant differences between the weighted method and the simpler method of using 0 and 1 for assigning scores to items were found. Phillips concludes that from the point of view of test construction, mathematically weighted scoring is probably not worth the time and effort.

Wilks (1938) reported a study in 1938 in which he presented a theoretical consideration of the problems of deriving score weights and concluded that in a long test of intercorrelated items, the method of weighting the individual items matters little.

Likert (1932) in his development of attitude scales, used procedures similar to those employed in ordinary test development. His scoring system was based upon the multiple choice method using three to five categories such as "Yes," "?," and "No" or five responses ranging from "Strongly Approve" to "Strongly Disapprove." He scaled the response categories using the category-scale method and used these scale values as weights for responses. He found, however, that values from one to five in the five choice items and two to four in the three choice items gave scores which were equally as reliable as the category scale values, and the two sets of scores correlated perfectly.

Strong reported a study in 1945 in which he used unit scales as substitutes for weighted scales in scoring the STVIB. Unit scale items were weighted 1, -1, or 0 instead of +4 to -4. He found that with small samples, weighted scales differentiated occupations better than did unit scales. However, as his criterion group enlarged, he found a corresponding decrease in the superiority of weighted over unit scales. Strong (1945).

Thus research evidence reported in the literature seems to support the use of arbitrarily assigned score weights in lieu of the more complicated mathematically derived weighted systems. Although the weights to be assigned items in this study were not to be empirically or mathematically derived, they were assigned weights ranging from four to one on the basis of ratings of each item's importance in contributing to a criterion as judged by a behavioral science panel.

III. Related Behavior Checklists

Although there are rating scales such as the Haggerty-Wickman-Olsen Rating Scale and the Rating Scale for Pupil Adjustment that are designed for the identification of children with behavior problems in a school setting; the writer has been unable to find a behavior checklist for this purpose in a search of the literature and a review of Buros' Mental Measurements Yearbook.

There are three checklists which are related to the instrument being developed in this study that warrant some attention. Newman developed a behavioral scale for assessing the learning and adjustment of six hyperactive males receiving therapy in a treatment center. Behavioral incidents were gathered on each subject daily which represented, ". . . complete pieces of behavior."

Newman (1960). Sample incidents from a total of seven hundred incidents were selected randomly for each subject. The selected incidents were built into a scale designed to evaluate a child's given behavior for each incident. All incidents used in the scale were submitted to a four member panel of judges to assess the degree to which incidents selected actually represented learning and adjustment behavior. An agreement index of .85 was reported by the author for the four judges in this task.

Weights for each item were assigned arbitrarily on the basis of an analysis of each incident's stimulus value for the subject. The instrument was used by Newman to assess behavior change for the six hyperactive males following treatment. Ratings from the behavior scale were divided into two parts for each child. The ratings from the first half of each subject's hospitalization were compared with those of the second half of the hospitalization period. Results indicated that behavior changes for five of the six children were significantly different from what could be expected by chance at the .001 level of confidence. Although this scale was not designed for use within a public school setting nor for the purpose of identifying children with behavior problems, the scale was reviewed because the author's experimental design is quite similar to that being used in the present study. Newman's procedure of collecting behavioral incidents is analogous to the process of abstracting operational descriptions of behavior from teachers' statements in the present study. Both designs use score weights assigned on a non-mathematical basis and both employ a panel of experts for the purpose of assessing each item's relationship to the behavior being measured. However, this instrument departs from Newman's scale in both purpose and basic orientation. It differs further in the sense that items were selected which could be directly observed by the classroom teacher and which did not require clinical judgment or the exercise of inference on the part of the observer.

Dreger (1964) is in the process of developing a behavior checklist which is designed for use by parents and/or teachers. The scale is as yet unpublished and is still in experimental form. The Behavior Checklist grows out of the Behavior Classification Project which began as an interdisciplinary attempt to develop a systematic classification of children's behavioral disorders. The rationale underlying the project was based upon the belief that the standard Psychiatric Association Nomenclature was not adequate for the purpose of classifying the behavior disorders of children.

The checklist was constructed from behavioral items that were, ". . . as purely descriptive of behavior as a team of experts and consultants could make and refine them." p. 2. Scale items were selected from many of the standard personality assessment instruments. In addition, fifty items were included in the scale which reflected parents' presenting complaints when they brought their children to mental health clinics. The final experimental form was developed by expanding the scale to include 229 items, submitting it to a panel of experimental psychologists and clinicians for criticism, and subsequent committee revisions from the project staff.

The checklist was subjected to a preliminary testing within thirteen child guidance clinics in 1961-62. Subjects described on the instrument were first admission children between the ages of six and thirteen. It was presented to parents by the card sort method with directions for sorters to note behaviors observed during the past six months and to include doubtful responses as negative responses. Usable records were obtained on 351 cases and were matched for age, sex, religion and socio-economic status with eight control subjects. Despite the presence of fifteen positive behaviors on the instrument, the number of yes responses was used as the criterion and the difference in this response between clinic and control groups was significant beyond the .001 level of confidence.

An inter-rater reliability check was performed on seventeen records of children from four additional clinics. For ten of the seventeen records the coefficient of agreement between the original parent sort and a later sort by another relative or close friend was .55, but the mean coefficient was .36. A later test-retest reliability check was reported which indicated an overall stability coefficient of .87.

Kvaraceus (1956) has developed an instrument called the KD Proneness Checklist which is designed for the identification of youth who are especially vulnerable to the development of delinquent behavior. The checklist contains a series of statements related to delinquent behavior such as "runs with a gang," "drunkenness in family," and "mother employed outside the home." Items are answered by an observer who checks "Yes," "No," or "?." Three research studies are reported by the author involving 130 delinquent boys and 434 boys and girls in a general school population. Results indicated that delinquents are usually given more checks in the yes column of the checklist. However, no studies are reported in which the observer was not aware that those being evaluated were delinquent.

Although the KD Proneness Checklist can be used in a school population, it has been designed for the detection of those especially vulnerable to becoming delinquent and not for identification of children who should be referred for psychological evaluation and/or treatment of behavior problems. However, the author seems to have made a genuine effort to include items in the checklist which are observable--a procedure that was duplicated in the present study. This procedure seems preferable to building scales or checklist for the measurement of internal feeling states in an effort to identify behavior problem children. If inferences are made about behavior on the basis of unobservable, internal emotional states, one must be able to validate their existence in order to make them acceptable. Reliably validating such emotional states represents an improbable, if not an impossible task.

Thus, it seems that the development of a behavior checklist which is composed of statements about overt, observable behavior and which is designed for use by the teacher would be of significant value in the identification of disturbed children and the referral of such children to appropriate treatment strategies. Such a scale would be useful in describing the actual classroom behavior of disturbed children and would be of value to psychologists in designing individualized treatment programs for children who are referred to them from classroom settings.

METHODOLOGY AND INSTRUMENTATION

I. Procedure for Collecting and Abstracting Item Pool Data

A population sample drawn from School District #4 in Eugene, Oregon, was chosen for this study. Research carried out on the Eugene school population shows a normal distribution on most educationally related variables. Socioeconomic surveys indicate a middle class population and school achievement of the population is high average when compared with national norms.

A random sample of thirty, experienced teachers was drawn from the population of fourth, fifth, and sixth grade teachers in the public schools of District #4. Each teacher in the sample was asked to identify those children in her class who exhibited chronic behavioral problems. Teachers were not provided with selection criteria but were instructed to simply identify behavior problem children in their classes. Teachers were then interviewed and asked to describe the nature of each child's problem, and to give operational descriptions of the behaviors that concerned them. Following Phillips' procedure of extracting increasingly refined levels of description, each interviewer asked the teacher specific questions as follows: "If I were to observe the child, what would I look for?" "You say he wants to try, how does he communicate this to you?" "In what way does he defy you?"

Interviews were typed and duplicated, and observable acts of behavior were abstracted from each interview, yielding an item pool of some three hundred items. Fifty of the most frequently identified behaviors from this sample were submitted to a panel of behavior scientists for an item rating task which is described below.

II. Behavior Science Panel Item Rating Task

The purpose of the panel's item rating task was to select appropriate score weights for assignment to behavioral items. Before discussing this task however, some mention should be made of the conditions which must be met in connection with the rating process in order to obtain meaningful results. Thorndike and Hagen (1961) point out that, "The best designed instrument cannot give good results if used under unsatisfactory rating conditions." p. 351. This statement has equal applicability to the rating of items as well as to ratees in relation to some criterion. Raters, for example, should be given detailed information on the type and kinds of judgments they are expected to make. Judges should have an intimate acquaintance with the material they are rating. A third condition for insuring optimum agreement among judges is the selection of raters with common qualifications, training, and interest in the subject matter field from which materials to be rated are lifted.

In the present study, an effort has been made to meet these conditions in an attempt to allow maximum agreement among judges to emerge. Five judges were employed in this study as opposed to a smaller number because of the greater reliability which emerges as the number of judges increases. For example, Thorndike and Hagen (1961) show that if the reliability coefficient obtained from one rater is represented by a r of .55, the reliability coefficients obtained by adding additional raters are as follows: two raters = .71; three raters = .79; five raters = .86; and ten raters = .92. p.363.

In this study, judges were asked to rate each item's effectiveness in contributing to a criterion on a twenty point scale ranging from of no importance to great importance. The scale is not divided into mutually exclusive categories but represents a continuum on which judges could rate an item at any given point. The form, with accompanying rating instructions for this task, is presented in Appendix A.

Judges' item ratings were pooled and averaged and each item assigned an arbitrary weight ranging from four to one on the basis of such ratings. Inter-judge reliability on the rating task was assessed by way of an analysis of variance technique which will be discussed further under methodology and procedure.

III. Data Collection

Items selected and weighted according to the above criteria were incorporated into a behavior checklist and given to a twenty-one teacher sample of fourth, fifth, and sixth grade elementary teachers. Teachers evaluated all pupils in their classes on the checklist after having observed them for approximately two months in the classroom situation. Each child evaluated on the instrument received a marking of either "Yes" or "No" for each item on the instrument which indicated the presence or absence of the behavior in question. Teachers were instructed not to single out problem children in their use of the checklist since such a procedure would have undoubtedly biased results. This procedure yielded scores on 534 fourth, fifth, and sixth grade children.

IV. Data Instrumentation

1. Reliability

A. Inter-Judge Reliability

The reliability of clinical judgment when more than two judges are used can be assessed by computing r_{11} for all possible pairs of judges and averaging them. The number of individual coefficients that would have to be computed is determined by $j(j - 1)/2$ where j is the number of judges. In this study, using five judges, there would be $5(5 - 1)/2$ or 10 possible coefficients to compute. This is a time consuming procedure and averaging correlation coefficients to obtain an overall measure of reliability is, at best, a risky process.

A more reliable and efficient method for estimating inter-judge reliability when more than two judges are used is by way of an analysis of variance technique using formula $r_{11} = \frac{MSs - MSe}{MS}$ where:

r_{11} = estimate of inter-judge reliability
MSs = mean square variance for subjects
MSe = mean square error

This technique is applicable regardless of number of points on the rating scale provided that such a scale is considered to be a continuum. The technique is inappropriate when the data are in discrete categories that cannot be ordered logically from least to most. It can be noted in Appendix A that the rating form chosen for this study meets all these requirements for applicability. Therefore, this technique was employed to assess inter-judge reliability in the present study.

B. Instrument Reliability

There are four procedures commonly used for testing the reliability of an instrument: (1) Test-retest (2) Alternate or parallel forms (3) Split-half technique (4) Rational equivalence. The method chosen for assessing reliability depends upon the purposes of the test, logistical requirements of the testing situation, and the type of instrument being developed.

In the split-half method, the instrument is divided into two equivalent halves. From the reliability of the half-test, the correlation of the whole test is then estimated by way of the Spearman-Brown or Kuder-Richardson prophecy formulas. In this procedure, two sets of scores are frequently established for correlational purposes by combining alternate items in the test. Thus, one set of scores would be made up of the odd-numbered items, 1, 3, 5, 7, and so on; while the second set of scores is comprised of even numbered items 2, 4, 6, 8, etc. Garrett (1962).

Garrett (1962) comments on the appropriateness of using the split-half method of measuring reliability by noting that,

> The split-half method is employed when it is not feasible to construct parallel forms of the test nor advisable to repeat the test itself. This method is regarded by many as the best of the methods for measuring test reliability. One of its main advantages is the fact that all data for computing reliability are obtained upon one occasion; so that variations brought about by differences between the two testing situations are eliminated. A marked disadvantage of the split-half technique lies in the fact that chance errors may affect scores on the two halves of the test in the same way, thus tending to make the reliability coefficient too high. However, the longer the test, the less the probability that effects of temporary and variable disturbances will be cumulative in one direction, and the more accurate the estimate of score reliability. p. 340.

A further advantage of the split-half method may result when the instrument being tested is designed to measure some aspect of personality or behavior. Thus the estimate of reliability is not affected by attitudinal/behavioral changes due to maturational factors or to other less predictable events when the split-half method is used.

Due to logistical limitations in the present study and due to the type of instrument being developed, it is most feasible to measure the reliability of the checklist by way of the Kuder-Richardson prophecy formula.

$$r_{11} = \frac{n}{n-1}\left[\frac{1 - M\left(1 - \frac{M}{n}\right)}{SD^2}\right]$$

where: M = the mean of the distribution of scores
SD = the standard deviation of the distribution
n = the number of items in the measure

C. Test Length Reliability

After a reliability coefficient has been computed for the checklist, the effect upon reliability of adding items to the instrument will be determined by way of formula:

$$rnn = \frac{nr_{11}}{1 + (n-1)\, r_{11}}$$

where: rnn = the correlation between n forms of a test and n alternate forms (or the mean of n forms versus the mean of n other forms)

If the reliability coefficient of the instrument were low with a length of fifty items, .60 or .70, for example, the coefficient may increase if additional items are added to the checklist. This formula will provide an estimate of the reliability coefficient increase obtained by doubling or tripling the length of the instrument.

2. Validity

An oft-cited criticism of checklists, rating scales, and inventories is that while a good deal of lip service is paid to the concepts of validity in test construction, no systematic effort is made to establish such validity empirically in the development of the instrument. In the present study, four types of validity were assessed: content validity, contrasted groups validity, criterion validity, and item validity. An item analysis was conducted to measure item variance and item discrimination value.

A. Content Validity

Provisions have been made in this study to insure that the instrument is composed of items which represent a sample of the behaviors which the checklist will measure. The checklist is composed of operational statements lifted from teachers' descriptions of problem behaviors of school children. The fifty most frequently identified, observable behaviors, determined by analysis of teacher descriptions, were selected for inclusion within the checklist. These fifty items were rated by five judges as to their importance in contributin to the study criterion.

B. Contrasted Groups Validity

In the contrasted groups method of assessing validity, two independent groups are defined in relation to the construct being measured and the instrument is then administered to both groups. Differences between the two groups in terms of test score are then tested for statistical significance. Levitt (1961) writing in Clinical Research Design and Analysis in the Behavioral Sciences explains the method by analogy. "Assume that it is known on some basis that population S contains a larger amount of the construct 'material' that we wish to define than does population T. Population S should therefore score higher on the average than population T on any index which is a valid measure of the construct. If this is found, by experimentation, to occur, then evidence for the validity of the operational definition may be reasonably claimed." p. 51.

In this study, the sample of 534 subjects evaluated with the instrument were screened for subjects who have been referred for psychological examination and treatment because of behavior problems observed within the classroom setting. Subjects were selected who qualified for any one of the following criteria:

(A) Has been examined by a psychologist and referred to a psychiatric or clinical facility (B) Specific educational provisions have been made for the subject within the school setting because of his behavior problem(s) (C) Has received instruction at home because of his inability to profit from classroom instruction due to his behavior problem(s).

On the basis of pilot information gathered in the present study and information provided by School District #4, it was anticipated that from forty to seventy children from the study sample would meet at least one of these criteria. Subjects thus identified were matched with subjects from the 534 pupil sample, not so identified, in terms of chronological age, sex, and grade in school. Subjects in both the criterion group and the matched controlled group were screened in terms of intelligence quotient; and the subjects with a reported intelligence quotient of 90 or below were excluded from the sample for purposes of this analysis. Differences between the two groups were tested by way of a t test of significance.

C. Criterion Validity

The procedure for determining the degree of relationship between the test score and the criterion in this study represents a special correlational problem. Garrett (1962) notes that, "In many problems, it is important to be able to compute the correlation between traits and other attributes when the members of the group can be measured in one variable but can be classified into only two categories in the second or dichotomous variable. When we can assume that the trait in which we have made a two-way split would be found to be continuous and normally disturbuted were more information available, we may compute a biserial r between the set of scores and two category groupings." p. 376. In this study, checklist scores were correlated with the criterion variable which was dichotomously divided into two groups: the criterion group and the matched control group. The criterion group was composed of those referred who qualified for any one of the three criteria discussed earlier. Members of the control group were matched with criterion group in chronological age, sex, and grade in school. It seemed reasonable to expect that those who have been referred to psychiatric or clinical facilities or those who require special educational provisions because of behavior problems should receive higher scores on the checklist than those who were judged not in need of such attention if the instrument measures problem behavior. The biserial correlation between the test score and the criterion was instrumental in answering such a question and provided an index of the instrument's predictive validity.

D. Item Analysis

1. Item Variance

In the present study, item variance indicies (pq) and item standard deviations (pq) were obtained for all fifty items. The item variance indices were computed by a formula reported by Guilford in *Psychometric Methods*. (1954). The variance of item I is given by the equation:

$$6^{2}i = PiQi$$

where: Pi = proportion passing the item or responding to the item in some specified manner

Qi = 1 - Pi

From this equation, the standard deviation of an item is

$$6i = \sqrt{PiQi}$$

The maximum variance of an item is obtained when fifty percent of the examinees pass the item or respond to it in some specified manner and fifty percent fail the item or respond to it in another specified manner. When p = .50 and q = .50, the item variance is .25 and the item is capable of making 2,500 discriminations among testees. If p = .70 and q = .30, the item variance is .21 and the item can make 2,100 separations among individual testees.

Garrett (1962) recommends item variance values of .24-.25 for most educational test items since it is desirable to make maximum separations among individuals in terms of mental ability, aptitude, and achievement factors by means of written tests. p.363.

However, when one is constructing an instrument which will separate a given or predetermined proportion of individuals from the total sample, the .24-.25 value for optimal selection of items does not apply. Lindquist (1950) notes that the purpose for which a test is to be used is more important in determining the number of separations which an item can make than are other considerations. If one wishes to discriminate between examinees capable of passing an item at the thirty percent level of difficulty and those not capable of doing so, then an instrument must incorporate an item of the thirty percent difficulty level. p. 309.

In the present study, it is important to be able to select items which are not so narrow or so limited in scope that they are useless for purposes of identification in that they very rarely occur within the classroom setting. A behavior such as physically attacks the teacher may occur frequently in a residential treatment facility for severely disturbed children but would probably occur very infrequently in the ordinary classroom setting. At the other end of the continuum, a behavior such as not paying attention is so common and so general that it is probably typical of most school children at one time or another. This behavior's innocuous content and high frequency would, in all likelihood, negate its utility and value in the identification process.

Numerous research studies have indicated that approximately ten percent of school age children possess behavior problems serious enough to require systematic treatment. Although estimates of the percentage of emotionally disturbed children in school populations vary considerably, ten percent seems to be the most frequently selected figure in discussions of this handicap in the literature. Kirk (1962). For purposes of identification, it becomes necessary to separate this ten percent, which makes up the disturbed population, from the rest of the school population. Therefore, it would not be feasible to select items with variance values of .24-.25 for inclusion in a scale for the identification of disturbed children.

In this study, a more appropriate criterion for item selection of the basis of variance indices would be from .09 to .16 since a value of .09 equals ten percent passing the item and ninety percent failing and a value of .16 is equal to twenty percent passing the item and eighty percent failing the item. (With reference to the scale being developed in this study, passing the item means possession of the behavior versus failing the item which refers to non-possession of the behavior.)

2. Item Validity

Guilford (1954) notes that the index of validity may mean how well the item measures or discriminates in agreement with the rest of the test or how well it predicts some external criterion. The most common indices used are pi, the proportion of examinees who pass the item, and either some measure of correlation of the item with an external criterion, r_{ic}, or the correlation of item with total score (internal criterion), r_{it}. He notes further that the correlation r_{ic}, of item with an external criterion, is less often computed and the intercorrelations of items, r_{ij}, are even less often computed. p.417.

In this study, a biserial correlation between scale items and the total score was computed, yielding a discrimination index which is a measure of internal consistency between individual items and test score. The specific procedure involved the selection of upper and lower groups, in terms of checklist score, according to Kelley's (1939) recommended criteria for the validation of test items, and then correlating each positive item with total score which served as the criterion variable.

The purpose of this procedure was to determine whether the two defined groups (upper twenty-seven percent and lower twenty-seven percent of the total sample) responded differently with respect to each item. The procedure determines the extent to which a given item discriminates among examinees who differ sharply in the function (behavior disturbance) which is measured by the test as a whole.

In summary, Guilford (1954) suggests that item analysis statistics are not just computed for their own sakes; but it is what one can do, knowing them, that is important. p. 417. For example, item analysis provides information, objective information, concerning the items that were written for the instrument. It provides an opportunity to check the writer's subjective judgment in selecting items to be incorporated into the instrument although it is no substitute for careful writing and editing of test items. An item whose validity index is .00 obviously does not contribute much to the instrument being developed. Through item analysis techniques, the test constructor is given an empirical base for accepting or rejecting items.

III. Educationally Related Variables

In a study of this type, it is important to find out what effect, if any, that non-behavioral variables have upon the obtained behavioral scores of the sample being evaluated. In this study, the instrument being developed is designed to measure behavior; yet it is conceivable that such educationally related but non-behavioral variables as grade of student, sex of student, and sex of rater could have an effect upon the checklist scores of the subjects in the study sample. Therefore, hypotheses have been constructed which are designed to provide a measure of the effect of such variables upon the scores of subjects in the sample.

IV. Hypotheses -- Stated in Null Form

Hypothesis one: The inter-judge reliability correlation coefficient will be 0.00.

Hypothesis two: The reliability correlation coefficient between split-halves of the instrument will be 0.00.

Hypothesis three: There will be no statistically significant differences in terms of checklist score between the criterion group and the matched control group.

Hypothesis four: The correlation between the criterion and checklist scores of subjects meeting criteria (A), (B), and/or (C) will be 0.00.

Hypothesis five: There will be no statistically significant differences between male and female subjects in terms of checklist score.

1. Sub Hypothesis A: There will be no statistically significant differences between male and female subjects in terms of checklist score in grades four, five, and six.

Hypothesis six: There will be no statistically significant differences between fourth, fifth, and sixth grade subjects in terms of checklist score.

Hypothesis seven: There will be no statistically significant differences between scores of subjects rated by a male rater and subjects rated by a female rater.

Hypothesis eight: There will be no statistically significant differences in obtained scores between subjects rated by a rater of the same sex and subjects rated by a rater of the opposite sex.

ANALYSIS AND DISCUSSION OF RESULTS

I. Reliability

A. Inter-judge Reliability

The purpose of the item rating task was to have five behavioral scientists rate the scale items on a continuum which ranged from zero to twenty. A zero rating indicated a behavior which is of no importance in handicapping behavioral adjustment and a rating of twenty designated a behavior which is of great importance in handicapping behavioral adjustment.

TABLE I

MEAN SCORES, STANDARD DEVIATIONS, AND INTER-RATER RELIABILITY (r_{11}) FOR ALL JUDGES ON FIFTY SCALE ITEMS

Judges	Mean	Standard Deviation
#1	11.8	4.1
#2	9.5	3.6
#3	9.5	4.4
#4	11.6	3.7
#5	12.7	3.5

Inter-judge reliability = .83

Since r_{11} equalled .83, the means of the five judges on all items were pooled and assigned as score weights for the differential weighting of the scale items. If r_{11} had not been acceptably large, there would have been no justification for using the item ratings of the five judges as differential score weights. Lindquist (1950) suggests that r_{11} = .60 is the minimum inter-rater reliability acceptable for this purpose.

TABLE 2

ITEMS RANKED IN DESCENDING ORDER ACCORDING TO MEAN RATING SCORE ON EACH ITEM BY FIVE JUDGES

Item	Mean Score
-Has no friends	16
-Has rapid mood shifts: depressed one moment, manic the next	16
-Utters nonsense syllables and babbles to himself.	15.4
-Other children act as if he were taboo or tainted.	15.2
-Repeats one idea, thought, or activity over and over.	15.2
-Does not initiate relationships with other children.	15
-Reacts to stressful situations or changes in routine with: body aches, head or stomach aches, nausea.	14.4
-Complains about others' unfairness and/or discrimination toward him.	14.4
-Expresses concern about something terrible or horrible happening to him.	14.2

TABLE 2--Continued

Item	Mean Score
-Has nervous tics: muscle twitching, eye blinking, nail biting, hand wringing.	14.2
-Complains of nightmares, bad dreams.	13.8
-Refers to himself as dumb, stupid, or incapable.	13.6
-Expresses concern about being lonely, unhappy.	13.6
-Shuns or avoids heterosexual activities.	13
-Is overactive, restless, and continually shifting body position.	12.8
-Makes distrustful or suspicious remarks about actions of others toward him.	12.8
-Doesn't protest when others hurt, tease, or criticize him.	12.8
-Perfectionistic: meticulous about having everything exactly right.	12.6
-Has temper tantrums.	12.2
-Disturbs other children: teasing, provoking fights, interrupting others.	12.2
-Comments that nobody likes him.	12.2
-Weeps or cries without provocation.	12
-Apologizes repeatedly for himself and/or his behavior.	12
-Does not engage in group activities.	12
-Has enuresis.	11.8
-Tries to avoid calling attention to himself.	11.8
-Is hypercritical of himself.	11.8
-Will destroy or take apart something he has made rather than show it or ask to have it displayed.	11.6
-Openly strikes back with angry behavior to teasing of other children.	11.6
-Habitually rejects the school experience through actions or comments.	11.6
-Displays physical aggression toward objects or persons.	11.4
-Frequently stares blankly into space and is unaware of his surroundings when doing so.	11.4
-Comments that no one understands him.	11.4
-Does not conform to limits on his own without control from others.	11.2
-When teased or irritated by other children, takes out his frustrations on another inappropriate person or thing.	11.2
-Is listless and continually tired.	11.2
-Has difficulty concentrating for any length of time.	11.2
-Reacts with defiance to instructions or commands.	10.8
-Becomes hysterical, upset, or angry when things do not go his way.	10.5
-Stutters, stammers, or blocks on saying words.	10
-Argues and must have the last word in verbal exchanges.	10
-Approaches new tasks and situations with an "I can't do it" response.	9.8

TABLE 2--Continued

Item	Mean Score
-Distorts the truth by making statements contrary to fact.	9.8
-Must have approval for tasks attempted or completed.	9.8
-Continually seeks attention.	9.6
-Underachieving: performs below his demonstrated ability level.	9
-Does not obey until threatened with punishment.	8.6
-Does not complete tasks attempted.	8.2
-Steals things from other children.	7.6
-Easily distracted away from the task at hand by ordinary classroom stimuli, i.e., minor movements of others, noises, etc.	6.5

TABLE 3

ITEM MEAN SCORES, CORRESPONDING SCORE WEIGHTS, AND NUMBER AND PERCENTAGE OF ITEMS IN EACH CATEGORY

Mean Score	Score Weight	N	%
16 15	4	6	12
14.4 13	3	8	16
12.8 12	2	10	20
11.8 6.5	1	26	52
		T = 50	100

In Table 3 it can be seen that six items or twelve percent of the total number of items were assigned score weights of four. Eight items or sixteen percent received score weights of three. Ten items or twenty percent received score weights of two, and twenty-six items or fifty-two percent received score weights of one.

With this weighting system, it would be possible for a given subject to receive a high score of one hundred and a low score of zero on the scale. In the 534 pupil sample used in this study, the highest score recorded was sixty-two, and the lowest score recorded was zero.

B. Instrument Reliability

In this study, the reliability of the scale was estimated by way of the Kuder-Richardson split-half method. The instrument was divided into equivalent split-halves by selecting odd and even numbered items for inclusion in the two half-tests.

In an effort to make the two halves of the test more nearly equivalent and to reduce the response bias which operates when a group of very deviant behaviors cluster together in serial form, items and their equivalent score weights were distributed equally among the two half tests. One behavior with a score weight of four was assigned as item number fifty and another behavior with a score weight of four was assigned as item number one. This procedure was duplicated for the remaining forty-eight items by alternately assigning score weights of four, then three, then two, and then one to the two halves of the scale.

The split-half reliability coefficient obtained on the scale was .98 with a standard deviation of 10.53 and a standard error of measurement of 1.28. A coefficient of .985 indicates that ninety-seven percent of the variance of test scores in the present sample is true-score variance and three percent of the test-score variance is error-variance. In terms of precision of measurement, the scale seems to be an excellent measure of true-score variance.

The correlation between a set of obtained scores and their corresponding true counterparts is given by the formula:

r loo = $\sqrt{r_{11}}$

where r 100 = the correlation of obtained and true scores

r_{11} = the reliability coefficient of the test

In this study, r loo = .98 which is the highest correlation this scale is capable of yielding in its present form. It is apparent from this analysis that revising or altering the scale in an effort to obtain a higher reliability coefficient would be impractical since it has already yielded the highest correlation coefficient of which it is capable. If the reliability coefficient had been .81, then the scale would have been capable of yielding an r of .90, and revision would have been more defensible.

With a reliability coefficient of .98, the scale is capable of making individual separations among subjects with a considerable degree of reliability as an r of .90 is the minimum coefficient acceptable for this purpose. In terms of reliability, the scale has met one of the major purposes for which it was designed--the separation of disturbed from non-disturbed school children.

C. Test Length Reliability

If the self-correlation of a test is judged unsatisfactory by the test constructor, he has the option of adding additional items to the test in an effort to increase this correlation coefficient. It should be noted that increasing the length of a scale n times in order to strengthen its reliability is no substitute for the careful construction of the original scale. Increasing the length of a poorly constructed test ten or fifteen times to improve its reliability represents an impractical solution to the problem of low reliability. Garrett (1962), p. 344.

In this study, formula $r_{nn} = \frac{nr_{11}}{1 + (n-1)r_{11}}$ was applied to the reliability coefficient in order to determine the effect upon the reliability of the scale by first doubling and then tripling its length. By this formula, a hundred item scale would yield an r of .99, and a hundred fifty item scale would also yield an r of .99. Thus the gain which would be realized by doubling or tripling the length of the scale would not be commensurate with the reliability increase which could be obtained with this procedure.

II. Validity

A. Contrasted Groups Validity

In the contrasted groups method of assessing validity, two independent groups are defined in relation to the construct being measured, and the instrument is then administered to both groups. Differences between the two groups in terms of test score are then tested for statistical significance. Levitt (1961).

In this study, two independent groups were defined in relation to the construct of behavior disturbance, and differences between them, in terms of checklist score, were tested for significance. Thirty-eight subjects in the 534 pupil sample were identified as behaviorally disturbed according to the criteria discussed earlier. Forty-six subjects in the sample qualified for one or more of these criteria, but eight were excluded from the experimental group since they had reported intelligence quotients of ninety or below. Although, it is recognized that many retardates have serious behavior problems, the purpose of excluding subjects with intelligence quotients of ninety or below was to separate, as nearly as possible, the effects of the construct of mental retardation from the effects of behavior disturbance which is the variable being measured in this study.

These thirty-eight subjects, so identified, were matched with thirty-eight subjects from the study sample, not so identified, in terms of age, grade, and sex. All subjects who matched the experimental S's in age, grade, and sex were lifted from the sample. A table of random numbers was applied to this group in order to facilitate the random selection of thirty-eight control subjects to be paired with the experimental subjects for purposes of experimental analysis.

TABLE 4

MEANS, STANDARD DEVIATIONS, AND N'S OF EXPERIMENTAL AND CONTROL GROUPS WITH TEST FOR STATISTICAL SIGNIFICANCE

Experimental (N=38)		Control (N=38)		D	Critical Ratio
M	S.D.	M	S.D.		
16.63	12.68	6.47	5.47	10.16	4.23***

* Significant at .05 level ** Significant at .01 level
*** Significant at .001 level

The difference between the means of the experimental and control subjects is significant beyond the .001 level of confidence. Contrasted groups validity can be reasonably claimed for the scale in that behaviorally disturbed subjects received significantly higher scores on the construct which the scale measures than did non-behaviorally disturbed subjects.

B. Criterion Validity

A biserial correlation was computed in this study to determine the degree of relationship between test score and the criterion (behavior disturbance). If the scale is measuring disturbed behavior, then it seems reasonable to expect that scores of subjects who have been referred to psychiatric or clinical facilities or those who require special educational provisions because of behavior problems should correlate higher with the criterion than scores of subjects who are judged not in need of such attention.

The biserial correlation between test score and the criterion yielded an rbis of .68. The standard error of this correlation is .039, and its index of predictive efficiency is .33. The rbis of .68 is significantly different from zero at the .01 level. The predictive efficiency index of .33 provides a measure of the scale's predictive value and indicates that the checklist has utilitarian value in the prediction of behavior disturbance in populations of elementary school children.

III. Item Analysis

A. Item Variance

TABLE 5

ITEM VARIANCE AND STANDARD DEVIATION INDICES
FOR FIFTY CHECKLIST ITEMS

Item	Variance Index	S.D.
1	.12	.69
2	.05	.47
3	.15	.78
4	.08	.58
5	.01	.25
6	.09	.60
7	.02	.29
8	.04	.43
9	.21	.92
10	.14	.78
11	.01	.28
12	.05	.50
13	.17	.85
14	.14	.76
15	.13	.74
16	.05	.48
17	.02	.33
18	.09	.63
19	.11	.67
20	.04	.45
21	.03	.39
22	.01	.22
23	.12	.33
24	.12	.70
25	.02	.28

TABLE 5--Continued

Item	Variance Index	S.D.
26	.02	.30
27	.04	.45
28	.03	.43
29	.09	.63
30	.04	.43
31	.03	.36
32	.05	.50
33	.00	.12
34	.01	.22
35	.12	.72
36	.00	.12
37	.06	.51
38	.13	.73
39	.07	.55
40	.05	.48
41	.17	.84
42	.08	.59
43	.04	.45
44	.01	.25
45	.12	.73
46	.04	.44
47	.00	.17
48	.03	.36
49	.21	.93
50	.10	.66

The range of item variance indices is from .00 to .21, and the item standard deviations range from a value of .12 to a value of .93. Seventeen of the items have variance indices which fall within the optimal range of .09 to .16 for the separation of the disturbed segment of the school population (approximately ten percent) from the remainder of the population. The remaining variance indices fall either slightly below or slightly above this range with the exception of items 33, 36, and 47.

The fifty items closely approximate the preselected standard of .09 to .16 chosen for judging the variance indices of the individual items. Items 33, 36, and 47 appear to be so narrow in scope as to be useless for purposes of identification. However, before rejecting these items on the basis of their item variance values alone, it would be useful to re-examine their values in a cross validation study conducted on another equivalent population sample.

TABLE 6

ITEM VALIDITY INDICES ON FIFTY CHECKLIST ITEMS

Item	Validity Index
1	.67 **
2	.19 **
3	.67 **
4	.65 **
5	.33 **
6	.09
7	.45 **
8	.42 **
9	.54 **
10	.61 **
11	.24 **
12	.49 **
13	.48 **
14	.65 **
15	.14 *
16	.55 **
17	.19 **
18	.59 **
19	.52 **
20	.38 **
21	.48 **
22	.12 *
23	.39 **
24	.56 **
25	.40 **
26	.35 **
27	.58 **
28	.48 **
29	.40 **
30	.57 **
31	.42 **
32	.60 **
33	.10
34	.26 **
35	.62 **
36	.10
37	.26 **
38	.55 **
39	.59 **
40	.30 **

TABLE 6--Continued

Item	Validity Index
41	.53 **
42	.12 *
43	.39 **
44	.15 **
45	.36 **
46	.59 **
47	.03
48	.15 **
49	.58 **
50	.32 **

* Significant at .05 level ** Significant at .01 level

The item validity indices on the fifty items vary from .03 to .67. In this analysis, total score was used as the criterion variable as opposed to an outside criterion which would have determined how well each item predicts that criterion. When total score is the dependent variable, the item validity indices are reflections of how consistently the individual items measure or discriminate in agreement with the total test.

The validity indices indicate that the individual items correlate highly with the criterion (total score) and that the items discriminate between subjects in the upper and lower twenty-seven percent of the sample in terms of checklist score. It should be noted that spurious correlation operates to inflate the individual item validity indices when total score is used as the criterion since each item constitutes a proportion of the criterion variable. Lindquist (1950), in discussing this problem, points out that there is no statistical technique by which the effect of the overlapping can be accurately removed with an increase in computational labor small enough to justify the resulting benefit. He suggests that the best that can be done is to indicate what the order of magnitude of the spurious correlation is likely to be and point out that the relative magnitudes of the item discrimination indices are affected less than their absolute magnitudes. p. 301.

The item validities suggest further that the items making up this scale constitute a very homogeneous, related set of behaviors with the exception of items 33, 36, and 47 which have validity indices of .10, .10, and .03 respectively. If these values were to remain constant or near constant in a cross validation study, then it would be incumbent upon the writer to either rewrite these three items or to reject them altogether in a revision of the scale. Other items in the scale with item validities below .20 would be treated in a similar fashion since items with validity indices of .20 and above are regarded as satisfactory. Garrett (1962) p. 301.

IV. Educationally Related Variables

Hypotheses were constructed in this study to determine the effect which non-behavioral but educationally relevant variables have upon the checklist scores of subjects in the study sample. These variables include, grade of student, sex of student, and sex of rater.

TABLE 7

SEX DIFFERENCES IN CHECKLIST SCORE ON ALL SUBJECTS

Male (N = 276)		Female (N = 258)			
M	S.D.	M	S.D.	D	Critical Ratio
10.50	12.16	4.83	7.40	5.67	6.67 **

* Significant at .05 level ** Significant at .01 level

TABLE 8

GRADE DIFFERENCES IN CHECKLIST SCORE ON ALL SUBJECTS

(N = 164) Grade 4		(N = 196) Grade 5		(N = 174) Grade 6				
M	S.D.	M	S.D.	M	S.D.	F Ratio	D	CR
						11.23**		
9.48	11.26	8.72	11.87				.76	.62
9.48	11.26			5.04	7.28		4.44	4.23**
		8.72	11.87	5.04	7.28		3.68	3.64**

* Significant at .05 level ** Significant at .01 level

TABLE 9

SCORE DIFFERENCES BY SEX OF RATER ON ALL SUBJECTS

Male Rater (N = 10)		Female Rater (N = 10)			
M	S.D.	M	S.D.	D	CR
7.12	10.53	8.43	10.39	1.31	1.47

* Significant at .05 level ** Significant at .01 level

TABLE 10

SCORE DIFFERENCES WHEN SUBJECTS ARE RATED BY A RATER OF THE SAME SEX VERSUS A RATER OF THE OPPOSITE SEX

Rating Comparisons	N	M	S.D.	F Ratio	D	CR
Male (R) rates Male (S)	148	9.60	12.80	17.67**	1.97	1.85
Female (R) rates Male (S)	127	11.57	11.04			
Male (R) rates Female (S)	128	4.26	7.41		1.72	1.89
Female (R) rates Female (S)	129	5.98	7.00			
Male (R) rates Male (S)	148	9.60	12.80			
Female (R) rates Female (S)	129	5.98	7.00		4.62	3.81**
Male (R) rates Male (S)	148	9.60	12.80			
Male (R) rates Female (S)	128	4.26	7.41		5.34	----**
Female (R) rates Female (S)	129	5.98	7.00			
Female (R) rates Male (S)	127	11.57	11.04		5.59	----**
Female (R) rates Male (S)	127	11.57	11.04			
Male (R) rates Female (S)	128	4.26	7.41		7.31	----**

* Significant at .05 level ** Significant at .01 level

TABLE 11

SEX DIFFERENCES ON ALL SUBJECTS BY GRADE

Grade of S	Male			Female			F Ratio	D	CR
	N	M	S.D.	N	M	S.D.			
Grade 4	87	12.02	13.63	77	6.62	9.00	14.25**	5.40	3.13**
Grade 5	102	12.63	14.03	94	4.47	6.92		8.16	----**
Grade 6	86	6.54	7.81	87	3.62	5.74		2.92	2.87**

* Significant at .05 level ** Significant at .01 level

Discussion

In Table 7, it can be seen that male students received significantly higher scores on the behavior checklist than female students. This result is consistent with research findings which have indicated that significantly higher proportions of boys than girls are identified as behaviorally disturbed. Beilin (1959). This finding also strengthens the applicability of the scale for use with school populations in that the checklist reflects sex differences which are known to exist in such populations in terms of behavior disturbance.

In Table 8, the analysis indicates that sixth grade students were rated as significantly less deviant than either fifth or fourth grade students. There is no empirical evidence, of which the writer is aware, that supports this finding. The result may be explained by the fact that the difference obtained represents a type one error in that no actual differences exist between the two groups even though the data appears to support the opposite conclusion. If this explanation were correct, then the null hypothesis would have to be accepted instead of rejected for this mean difference. Since the critical ratios between both fourth and sixth and fifth and sixth grade subjects was significant beyond the .01 level, this explanation is possible but highly improbable. Another explanation may be that sixth grade students are rated as less deviant than fourth and fifth grade students because of some as yet unexplained and unresearched maturational processes. A third possible explanation may be that the teachers who rated sixth grade students in this study were "easier" raters than fourth and fifth grade teachers. All three of these possible explanations are speculative and would be very difficult to test experimentally.

No statistically significant differences were found between male and female raters on their ratings of all subjects. This result indicates, as would be expected, that male raters did not rate subjects as significantly more or less deviant than female raters. This would suggest that the male teachers in this study are not "harder" or "easier" raters than female teachers.

An analysis of variance applied to the means of subjects rated by a rater of the same sex and subjects rated by a rater of the opposite sex yielded an F ratio which was significant beyond the .01 level. However, inspection of the respective means indicates that male and female raters do not rate male subjects in a significantly different fashion; nor do male and female raters rate female subjects in a significantly different fashion. Thus, a same sex bias does not appear to be operating in the ratings of teachers in this sample. The major part of the variance is accounted for by the fact that both male and female teachers rated male students as significantly more deviant than female students.

The analysis in Table 11 for sex differences across grades four, five, and six yielded an F ratio which is significant beyond the .01 level. Inspection of the means reveals that sex differences between male and female subjects in terms of checklist score, held constant across the three grades. It should be noted that even though sixth grade subjects were rated as significantly less deviant than fourth and fifth grade subjects, sex differences between male and female subjects in grade six were statistically significant.

CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

Review of Null Hypotheses

Hypothesis one stated that the inter-judge reliability coefficient in this study would be 0.00. The obtained coefficient, as determined by an analysis of variance technique, was .83 which requires the rejection of the null hypothesis at the .01 level. This measure of inter-judge agreement provided justification for using the pooled mean scores of all judges as differential score weights for the individual scale items.

Hypothesis two, which stated that the reliability correlation coefficient between split halves of the test would be zero, was rejected at the .01 level. The actual reliability coefficient for the scale was .985 which indicates that the scale possesses considerable internal consistency and that .97 percent of the total variance is accounted for by the fluctuation of true scores as opposed to .03 percent of the total variance which is accounted for by error variance.

Hypothesis three postualted that there would be no statistically significant differences, in terms of checklist score, between the criterion group and the matched control group. The null hypothesis must be rejected for this analysis as the mean score difference between these two groups was significant beyond the .001 level. The scale thus appears to be capable of discriminating effectively between these two populations, and it possesses contrasted groups validity to the extent measured by a probability value of .001.

Hypothesis four predicted that the correlation between the criterion and checklist scores of subjects meeting criteria (A), (B), and/or (C) would be 0.00. The biserial correlation coefficient computed for this analysis yielded an rbis of .68 which is significantly different from zero beyond the .01 level. The null hypothesis must therefore be rejected in this analysis. The correlation suggests that there is a considerable relationship between high scores on the checklist and the construct of behavior disturbance. The predictive efficiency index indicates that the scale is capable of predicting this construct to an extent expressed by a value of .33.

Hypothesis five, which stated that there would be no statistically significant differences between male and female subjects in terms of checklist score, was rejected at the .01 level. Male students were rated by teachers in the sample as significantly more deviant than female students on the behavior checklist. This result is consistent with the ratio of males to females who were identified as behaviorally disturbed in the study sample. Of the original forty-six subjects, who were identified, thirty-four were males and twelve were females. The null hypothesis for sub hypothesis A was rejected at the .01 level since these sex differences remained constant across the three grades.

Hypothesis six predicted that no statistically significant differences would exist between fourth, fifth, and sixth grade subjects in terms of checklist score. The null hypothesis was rejected since sixth grade students were rated as significantly less deviant than either fifth or fourth grade students. The critical ratio for this difference was significant beyond the .01 level.

Hypothesis seven stated that there would be no statistically significant differences between scores of subjects rated by a male rater and subjects rated by a female rater. The null hypothesis was accepted for this analysis since the differencesbetween the means of students rated by male teachers and students rated by female teachers were not statistically significant. This result indicates that male and female teachers did not rate students in a significantly different fashion in this study.

Hypothesis eight predicted that there would be no statistically significant differences in obtained scores between subjects rated by a rater of the same sex and subjects rated by a rater of the opposite sex. The null hypothesis was rejected for this manipulation since an analysis of variance yielded an F ratio of 17.67 which was significant well beyond the .01 level. However there were no significant differences between the means of male and female teachers who rated male students and male and female teachers who rated female students.

The behavior checklist developed in this study does appear to have relevant applicability for the task of identifying behaviorally disturbed children within school populations. The validity of the scale, as determined experimentally, indicates that the instrument is measuring the construct which it was designed to measure--behavior disturbance. The reliability coefficient suggests that it measures this construct in an internally consistent fashion. The stability of its measurement function, however, must be determined by a test re-test measure of reliability. With the exceptions of items 33, 36, and 47, the individual behaviors included in the checklist appear to be suitable for the purpose of measuring behavior disturbance. It is hoped that the scale will facilitate the identification of behaviorally disturbed children in school populations and that it will prove useful to psychological personnel in designing treatment programs for disturbed children who are referred to them from school settings.

Implications for Further Research

The implications which the development of this scale has for further research are evident in the areas of cross validation, normative sampling, concurrent validation, test re-test reliability measurement, and multiple ratings of the same student by different teachers.

Before any extensive conclusions are drawn about the applicability of this scale to school populations in general, the scale should be cross validated on one or more samples which are comparable to the sample used in the present study. In such a research project, it would be important to determine whether the validity and reliability results obtained in this study hold constant in other, equivalent samples.

If the scale is going to be used on any kind of regional basis, it would be important to establish norms for the given age, grade, and sex distributions of school children. If the sampling process were adequate, appropriate cut-off points could be established for such distributions and school personnel using the scale would be in a better position to make meaningful separations and referrals among school children in terms of placement of such children in existing treatment programs.

There are a number of behavior checklists, designed for the identification of disturbed children which are being currently developed in research projects across the country. It would be useful to concurrently validate this scale against one of these checklists in order to compare them in terms of their consistency and accuracy in the measurement process.

Since a major portion of the low reliability reported in teachers' identification of disturbed children has been attributed to inter- and intra-teacher variability in their judgments of such children, there is a need to match a number of teachers on such variables as age, sex, years of teaching experience and have them describe the same child on an appropriate measuring instrument. Such a project presupposes that two or three matched teachers would have observed the subject being rated for equivalent amounts of time. The results of such a study would be very useful in providing information about the variability among teachers in their ratings of disturbed children. An instrument like the one developed in this study could serve as one of the instruments used in this process.

Any further research conducted on the behavior checklist developed in this study should include a test-retest measure of reliability in order to assess the stability with which the scale measures disturbed behavior. Garrett (1962) has so aptly pointed out that chance errors tend to become cumulative in one direction when the split-half estimate of reliability is used. Therefore, it would be incumbent upon the writer to obtain a test-retest measure of reliability on the scale as a basis for comparison before releasing the scale for systematic or extended use.

SCALE DEVELOPMENT PROCEDURES

SECTION II

BEHAVIOR RATING SCALE

I. Methodology

1. Scale Construction

An item pool of 189 observable statements about behavior were submitted to a panel of behavioral scientists for the purpose of developing an educationally relevant behavior classification system. After construction of the behavior classification system, these items or behavioral statements were refined and incorporated into an appraisal instrument which was designed for the purpose of estimating the prevalence of social-emotional problems among fourth, fifth, and sixth grade children within the Eugene School District.

The items were devoid of traditionally used psychological terminology and reflected the major concerns which teachers have in their interactions with students in their classes. This latter assumption is supported by the process that was used to collect the item pool data, i.e., teachers were asked to describe the behaviors of disturbed children in terms of the extent to which they disrupted their classes or created generalized disturbances within the school setting.

It has been argued that a major cause of teacher/psychologist disagreement in the identification of disturbed children is that the teacher's role, teaching, is quite different from the psychologist's role, treating. Trippe (1961). Thus teacher's emphasize behaviors which are disruptive of class order while psychologists emphasize behaviors which impair the child's social/behavioral functioning. The purpose of the sorting task was to help bridge the artificial dichotomy which exists between the roles of teaching and treating and to strengthen the degree of relationship between identification and treatment criteria.

The panel of behavioral scientists was asked to sort these 189 behaviors into educationally relevant behavioral categories of their own choosing. The panel was composed of: a school psychologist, a remedial teacher, a social worker, a psychologist, and a child psychiatrist. The panel sorted the represented dimensions of behavior into categories which were, in their estimation, comprehensible to members of their own professions. The expected outcome of the sorting task was a behavior classification of the scale items that would be educationally prescriptive and which would facilitate treatment decisions and referrals by psychological personnel in the school setting.

After the panel members had independently sorted the behaviors along dimensions of their own choosing, the established behavioral categories were refined into an eight category behavioral classification that was acceptable and functional for all panel members. This system accounted for 124 of the 189 behavioral items. The remaining 65 items were judged as either educationally irrelevant or inappropriate for this particular classification of behavior. Items representing measures of acting out, disruptive behaviors were randomly assigned to the first section of the scale. Items which provide measures of restricted functioning and withdrawal behaviors were randomly assigned to the second section of the scale. In the scoring section, item scores for each item are assigned to their behavioral categories. A sub-group score is thus obtained for each behavioral category. These component scores are then transformed into a composite score for each subject.

In Section One of the scale, three response measures are obtained on each item. These measures are: (1) Rate of Occurrence, (2) Rater Response, (3) Rater Reaction. Rate of occurrence provides a measure of the frequency with which a given behavior is emitted over time. Rater response determines how the teacher (or rater) responds to different behaviors as they occur within the classroom setting. This measure is designed to secure data on whether the teacher's typical response operates to reinforce or extinguish deviant classroom behavior. Rater reaction indicates the extent to which a given teacher is disturbed or irritated by deviant behaviors emitted within the classroom setting. A tentative hypothesis has been developed in the current project which argues that deviant behaviors which are highly irritating or disturbing to the ordinary classroom teacher are significantly more predictive of an educational or psychological referral than are equally handicapping deviant behaviors which are less disturbing for the teacher.

2. Initial Testing and Results

1. Identification/Selection Sample

In the process of identifying subjects for inclusion within an experimental class for disturbed children, initial data was collected on a sample of seventeen subjects and raters. Preliminary analysis of the data has yielded the following results.

TABLE I

MEANS AND SIGMAS FOR SECTION I (ITEMS 1-64) SECTION II (ITEMS 65-124) AND TOTAL (ITEMS 1-124)

Section I		Section II		Total	
$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.
87.35	38.74	79.05	37.69	83.20	38.21

TABLE II

MEANS AND SIGMAS OF INDIVIDUAL RESPONSE MEASURES

Rate of Occurrence		Rater Response		Rater Reaction	
$\bar{x}$	S.D.	$\bar{x}$	S.D.	$\bar{x}$	S.D.
87.35**	38.74**	73.88	29.10	71.35	29.18

*Significant at .05 level ** Significant at .01 level

TABLE III

CORRELATIONAL INDICES BETWEEN SCALE SECTION I AND SECTION II AND BETWEEN RESPONSE MEASURES I, II, AND III

Sections I/II	Rm I/II	Rm II/III	Rm I/III
r = .84**	.93**	.88**	.85**

* Significant at .05 ** Significant at .01

The mean score for the disturbed children included in the present sample was 83.20 with a sigma of 38.21. Mean scores would indicate that the subjects sampled received higher and more frequent scores on the scale section representing disruptive, acting out behavior than they did on the section which measures withdrawn, restricted behavioral functioning. This was a predictable outcome in that the selection procedures were biased toward isolating and identifying acting out, disruptive subjects for inclusion within an experimental setting.

Results of the three item response measures suggest that there is less variability across teachers in their responses and reactions to emitted deviant behaviors than in their judgments about the current status of these behaviors. Teachers in the present sample also responded to emitted behaviors (Rm_1) with approximately the same frequencies on response measures two and three. It is not possible, in this analysis, to determine whether this result represents a constant rating error by teachers or whether it approximates an existing condition in the educational environment. The scale will be submitted to a more representative sample of 100 teachers for further testing during the academic year 1967-68. The resulting data will be subjected to a more intensive analysis and verification process at that time.

The correlation between sections one and two expresses the degree of relationship which obtains between scores on acting out and withdrawn item measures in the same subject. The obtained correlation on seventeen subjects was .84 between these two behavioral dimensions. This result would indicate that the presence of acting out and withdrawal behaviors, as defined by the scale items, are not incompatible within the same subject.

The correlations between response measures I and II, II and III, and I and III were .93, .88, and .85 respectively. The relationship between variables I and II indicates that the teacher responds with more intensive aversive controls as the frequency of the behavior increases. The correlation between I and III suggests that there is an inverse relationship between the teacher's tolerance level for emitted deviant behavior and the frequency of that behavior. As the behavior increases in frequency, the teacher's tolerance (as measured by a disturbance index) for that behavior correspondingly decreases. This result would seem to suggest that teachers react more differentially to the frequency of any deviant behavior rather than to the specific type or class of deviant behavior.

2. Experimental Class Testing and Results

Six subjects, who were members of an experimental class for the treatment of disturbed children, were rated on the scale by three judges who continuously observed their behavior for a minimum of two hours per day. The subjects were male children in grades four, five, and six, who were undergoing treatment in the current research project. The judges observed each child for a period of two weeks before making their initial ratings (Rt_1). After a period of six weeks, the judges were asked to rate the same subjects a second time (Rt_2). The judges were instructed to rate the current status of the behavior on each rating session. The purpose of these instructions was to allow changes in the status of the behavior, as a function of treatment and SV factors, to emerge between Rt_1 and Rt_2. Accordingly, test-retest measures of the stability of the judges' ratings are relatively meaningless in this application of the scale. Results of this application are presented below.

A. Inter-rater Reliability

Figure I

GRAPH I

Inter-rater Reliability Coefficients
For Individual Items (Total N = 124)

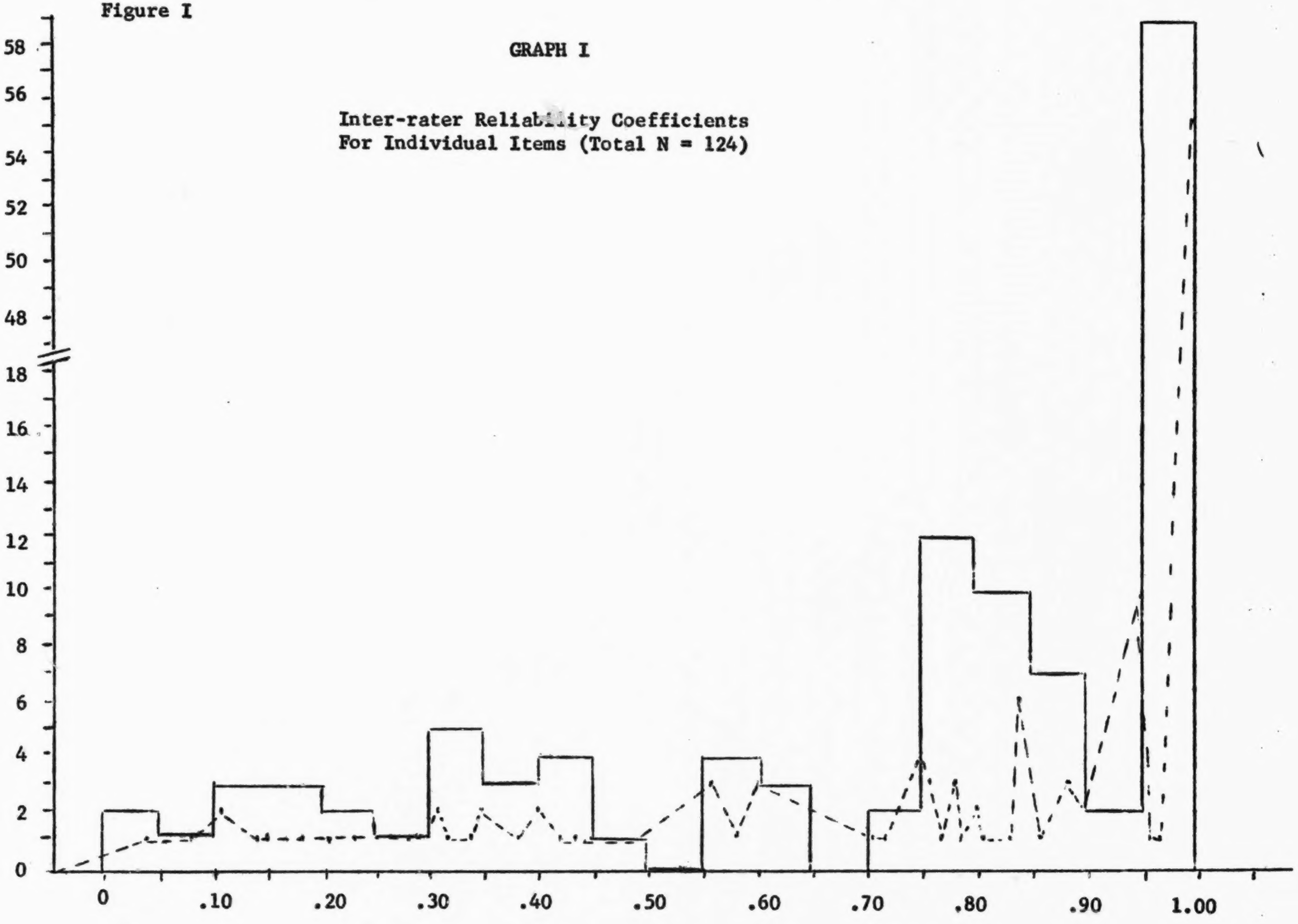
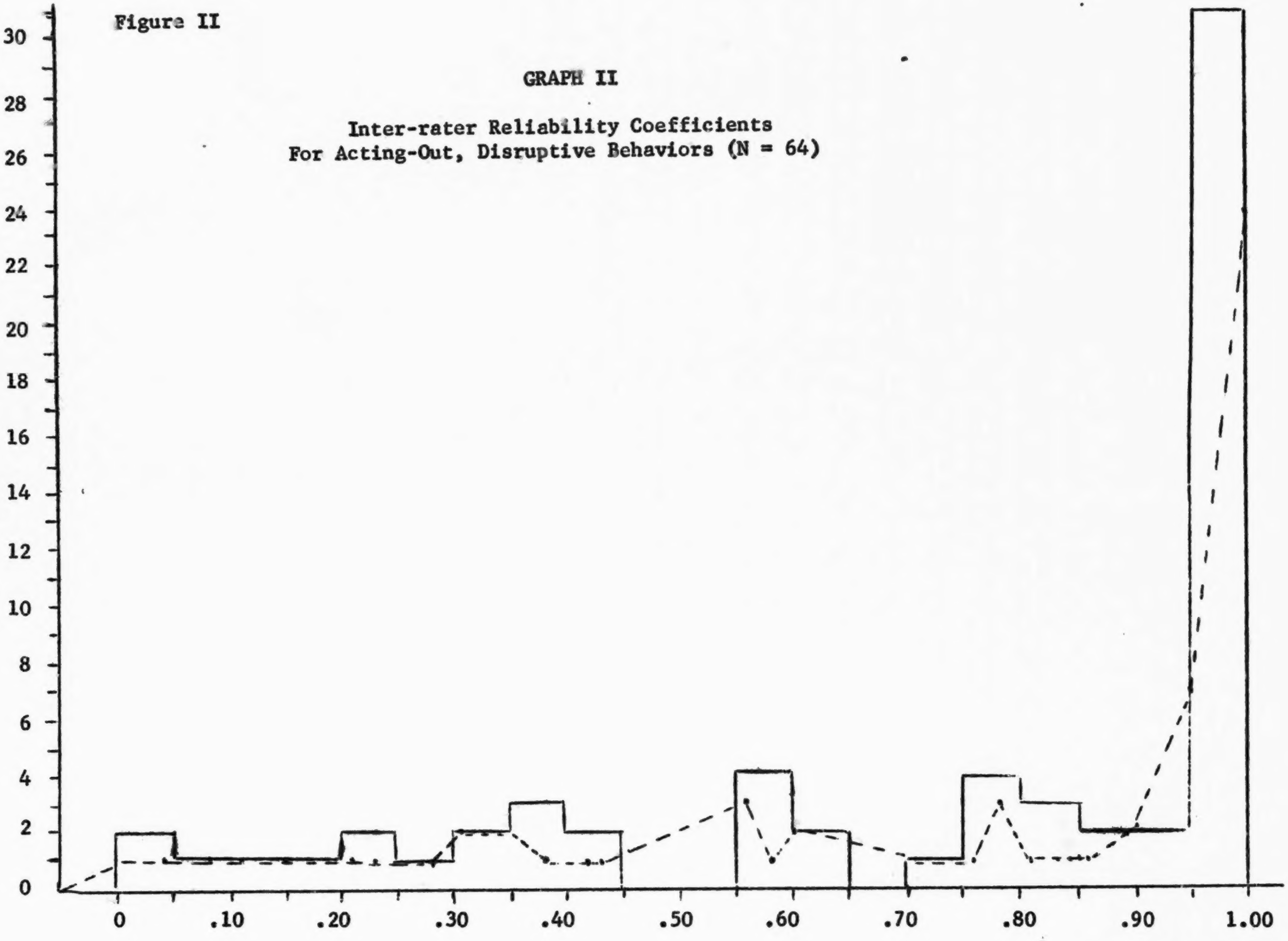


Figure II



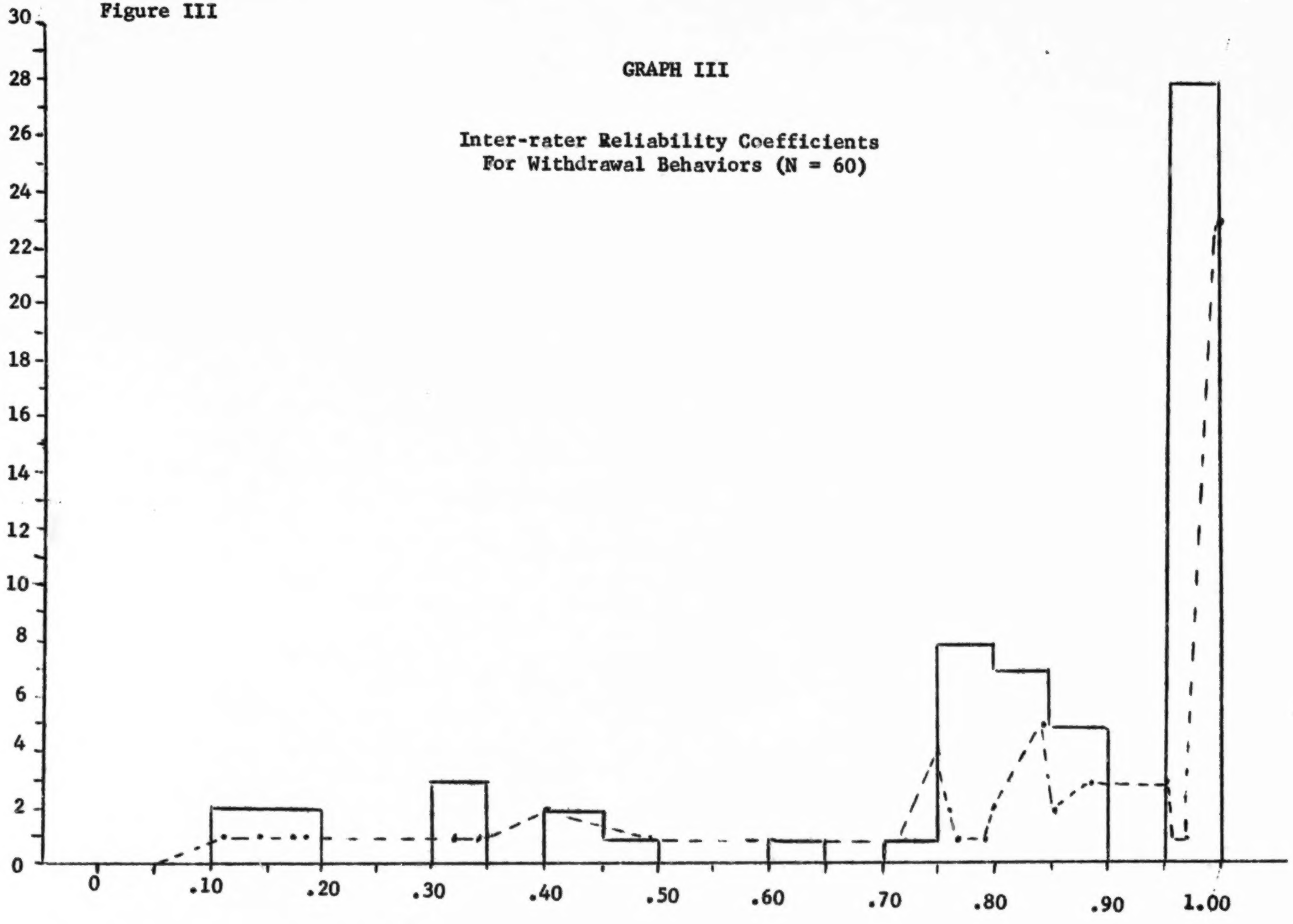
GRAPH II

Inter-rater Reliability Coefficients
For Acting-Out, Disruptive Behaviors (N = 64)

Figure III

GRAPH III

Inter-rater Reliability Coefficients
For Withdrawal Behaviors (N = 60)



The analysis of the extent to which judges agreed in their ratings of the frequency with which a behavior occurs within a given subject is graphically depicted in Figures I, II, and III. By way of conversion to Fisher's Z function, the average inter-rater reliability coefficient for all 124 scale items was .935. The mean value for acting out, disruptive items was .93 and the mean value for withdrawal behaviors was .94. The mean difference of .01 was, of course, not statistically significant.

B. Treatment Differences

TABLE IV

WILCOXON MATCHED PAIRS SIGNED-RANKS TEST FOR DIFFERENCES WHEN SUBJECTS ARE USED AS THEIR OWN CONTROLS

Subject	Rt_1	Rt_2	d	Rank of d	Rank w/LFS
1	99	76	23	5	--
2	67	59	8	2	--
3	74	73	1	1	--
4	83	72	11	3	--
5	95	74	21	4	--
					T = 0

H_0 rejected in favor of H_1 -- $p = .01$ ($N = 5$)

TABLE V

WILCOXON MATCHED PAIRS, SIGNED-RANKS TEST FOR DIRECTION AND SOURCE OF BEHAVIOR CHANGE WHEN SUBJECTS ARE USED AS THEIR OWN CONTROLS

	Subjects	Rt_1	Rt_2	d	Rank of d	Rank w/LFS
Acting Out Behaviors	S_1	64	43	21	10	
	S_2	44	36	8	8	
	S_3	47	42	5	4.5	
	S_4	52	45	7	7	
	S_5	51	45	6	6	
Withdrawal Behaviors	S_1	33	36	3	2.5	
	S_2	22	22	0	1	
	S_3	30	27	-3	-2.5	2.5
	S_4	31	26	-5	-4.5	4.5
	S_5	41	28	-13	-9	9.0
						T = 16.0

H_0 accepted

The rating scale was applied to the experimental class subjects in order to determine whether the scale ratings would reflect treatment differences which were known to exist. The results in Table IV indicate that the scale did reflect behavior changes in the experimental class subjects. H_0 stated that there would be no differences between pre- and post-behavior ratings when subjects are used as their own controls. H_0 was rejected in favor of H_1 at the .01 level.

Since the treatment model represented a therapeutic as opposed to a prosthetic application of learning theory principles, it was hypothesized that the major source of behavior change would occur in acting out, disruptive behaviors instead of in withdrawal behaviors. This hypothesis was not supported by the data in Table V. H_0 was accepted in this analysis.

Discussion

The initial data collected on the scale has been drawn from small samples that are less than representative of a given population. Therefore, conclusions drawn from analyses of the data are regarded as tentative and speculative. During the academic year 1967-68, procedures will be implemented which are designed to estimate the reliability of the scale and to begin the task of establishing its validity. These procedures were included in the Project Status Report (February, 1967) and will not be discussed here. It should be noted, however, that the investigators plan to investigate the teacher variable in the validation process in terms of its functional relationship to behavior disturbance in children.

SCALE DEVELOPMENT PROCEDURES

SECTION III

BEHAVIOR OBSERVATION FORM

Behavioral Observation Form

The third stage of assessment in the identification model represents a time sampling technique which measures task-oriented behavior by way of a behavioral observation form. This sampling technique serves several purposes in the current project. It is used to verify the judgments of teachers' ratings of disturbed behavior. It functions as a criterion measure for behavior change in experimental class subjects as a result of treatment. It is the most sensitive and reliable measure of the status of behavior in the model, and it therefore carries more weight in determining whether a given child is referred for the treatment process.

Independent observers use the observation form to collect time samples of behavior during three different phases in the treatment cycle. Data is collected on subjects during a pre-treatment phase in the regular educational setting, during the treatment process, and during a follow-up observation period when subjects are returned to the regular educational setting.

Each observation session represents a period of ten minutes. This time sample is divided into sixty ten second intervals. Observers are required to record the behavior as it occurs in each ten second interval. In its current form, there are five possible behaviors that an observer could record on the behavioral observation form. These are: (1) TOI = Task oriented - independent, (2) TOD = Task oriented - dependent, (3) NTD = Non task deviancy, (4) H = Hand raising, and (5) D = Distraction. (An explanation of these behavioral categories is contained in Appendix D.) Ratio's can be computed for these five behavioral variables which yields data on the type of task-oriented or non-task-oriented behavior that a given subject emits. This form has proven to be a very sensitive measure of task-oriented and non-deviant behavior in the current project, since deviant behavior is incompatible with task-oriented behavior as defined by the observation form. The form is easily modified and revised, and it is expected that it will be further refined as the project develops.

TREATMENT EFFECTS

SECTION IV

THE DEVELOPMENT OF EDUCATIONAL PROCEDURES FOR USE WITH BEHAVIORALLY DISTURBED CHILDREN

THE DEVELOPMENT OF EDUCATIONAL PROCEDURES FOR USE WITH BEHAVIORALLY DISTURBED CHILDREN

The academic and social behaviors of children who function productively in a regular classroom setting are ordinarily under the control of a wide variety of generalized reinforcers natural to that setting. Solving problems, completing assignments and success at academic endeavors in general function as powerful reinforcing events which maintain academic behavior. Such behaviors are further strengthened as a result of the parental and teacher administered praise that often accompanies appropriate academic behavior.

It is not surprising, then, that the behaviors of most school children are responsive to traditional educational procedures and methodologies even where no systematic efforts are directed toward gaining behavioral control. The "acting out" child, however, complete with accompanying academic disabilities, often misses out on these avenues of positive reinforcement natural to the setting. Reinforcements for academic behavior are rarely available. The low probability of success and/or praise being associated with academic behavior decreases the frequency of academic behavior in a spiraling process, i.e., the fewer the reinforcements, the less academic work attempted; the less work attempted, the fewer the reinforcements. In addition, many of the social behaviors demonstrated by these children are aversive and thereby preclude or severely limit the probability of the child being positively reinforced by teachers or peers. Social approval or praise often has little desired effect on these children. In fact, there is some evidence (Johns and Quay, 1962; Levin and Simmons, 1962) which suggests that adult praise is aversive for "acting-out" children.

Early attempts to treat the behaviorally disordered child in special classes within the school setting met with little demonstrable success. Kounin, Friesen, and Norton (1966); Rabinovich (1959); and Shannon (1961), suggest that the inability of the schools to deal effectively with these children stems primarily from the unavailability of established procedures and techniques that might be effectively employed within the context of the regular school setting.

The experimental analysis of behavior undertaken by Skinner (1938) revealed many principles from which are derived valuable behavior modification techniques. The success of these techniques in changing behavior has been widely demonstrated in laboratory settings.

Recent extensions of these same principles to the behavior of deviant children in applied settings have also met with considerable success (Patterson, 1965 (a) (b); Stranghan, 1964; Zimmerman and Zimmerman, 1962). These studies reflect, for the most part, behavior modification with individuals or small groups in highly controlled settings. The feasibility of adapting behavior modification techniques for use in the regular school setting by regular school personnel remains undemonstrated. Quay, et. al. (1966) emphasized the importance of extending these principles to the "grass roots" level in their suggestion that:

> "The economics of public schools obviously require the development of techniques that will allow children to be handled in a group situation by as few adults as possible. Most of the techniques of behavioral remediation have been developed for use on an individual basis and it seems crucial at this stage to attempt to extend these techniques to group situations. . . . Behavior techniques . . . are likely to remain economically unfeasible, unless they can be adapted for use in a group setting such as the classroom."

One such adaptation of these behavioral principles to group settings is the token economy system which has often proved successful where traditional educational procedures have failed. (Girardeau and Spradlin, 1964; Birnbrauer and Lawler, 1964; Quay, Werry, McQueen, and Sprague, 1966). The token reinforcers may be tangible or symbolic. Their value is derived from the various kinds of "back-up" reinforcers (candy, trinkets, free time, etc.) for which they are exchanged.

Once the desired behaviors come under stimulus control, the less "natural" back-ups reinforcers are gradually eliminated and replaced by reinforcing stimuli more readily available in the natural environment. This process is accomplished by pairing the presentation of the "artificial" reinforcer with a more natural reinforcer and gradually "fading" the presentation of the less natural reinforcer.

The schedules of reinforcement employed in the initial stages of behavioral acquisition are often atypical of those present in the natural setting. In the initial stages of acquiring a behavior, it is often necessary to reinforce on a continuous or small ratio reinforcement schedule. The child receives great quantities of reinforcement for minimal production. Once the behavior comes under control, however, the schedules of reinforcement are gradually increased so that the child is responding at high levels for minimal reinforcement.

A major objective of this project is to develop a set of general strategies and specific methodologies that will enable school personnel to efficiently meet the educational requirements of behaviorally disturbed children within the context of the regular school setting. The following sections describe the procedures and results obtained with the first two groups of acting-out children enrolled in the experimental classroom.

Subjects - Group I

The first group of students enrolled in the experimental classroom consisted of five fifth and sixth grade boys. The two major selection criteria were average or above intellectual ability and a demonstrated chronic failure to progress academically and socially within the regular classroom. Each child evidenced a number of behaviors that made him a poor candidate for learning. Physical and verbal abuse of teacher and peers, hyperactivity, defiance and other behaviors generally incompatible with academic pursuits were prevalent to a high degree. Task oriented academic behavior was conspicuous by its absence. One boy was permanently removed from regular class placement due to the unavailability of effective controls for his violent acting-out behavior.

Setting

The classroom was located in one of the elementary schools in the participating school district. All students enrolled in the experimental classroom, including those from other elementary schools in the district, were enrolled in regular classes at this school.

The physical arrangement of the experimental classroom included the student desk area where academic assignments were undertaken, a series of tables located along two walls where leisure reading, science, art, and music materials were provided and two high interest rooms with sink facilities for science experiments, crafts, and model building. A "time-out" or isolation room equipped with desk and chair adjoined the main classroom.

Procedures - Group I

The class was operated on a half-day basis, leaving the afternoons available for the children to return to the regular class. This approach of combining special and regular class placement into one program was believed to have several distinct therapeutic advantages. It allowed for the integration of the behaviorally disordered child with his "normal" peers. It also provided a learning situation that could be individually tailored to meet each child's specific academic and social requirements. Initially some children, unable to function within the regular classroom, were "full time" students in the experimental class. Other children with less deviant behavior were able to operate on the half day regular class and half day experimental class basis. This administrative structure provided an opportunity for a gradual return of the child to the regular classroom as his academic and social behaviors came under the control of the response-reinforcement contingencies in operation in the regular classroom environment. Such an arrangement facilitated communication between the project staff and the regular classroom teacher. An attempt was made to adapt strategies and techniques developed within the framework of the experimental class for use in the regular class.

One aspect of the reinforcing climate established within the experimental classroom consisted of "free time" to engage in high interest activities which the students earned by demonstrating appropriate academic and social behavior. The use of one behavior to reinforce or increase the probability of another behavior is an adaptation of the Premack (1959) principle. Simply stated, the Premack principle means that any behavior is strengthened or will increase in probability of occurrence when followed by a behavior which occurs at a high independent rate.

Observation of the activities of the students enrolled in the class revealed that academic task oriented behavior was a low frequency behavior and that building model airplanes, cars, and craft objects occurred at a high rate. Free time, the opportunity to engage in a variety of high frequency behaviors, was made contingent upon productive academic as well as social behavior.

Free time was selected as the primary reinforcing event for several reasons. First, it allowed each individual child to choose the free time activity that was most reinforcing to him. Furthermore, the dimension of time can be readily broken down into small units, which makes it an ideal reinforcer. Free time also has an advantage over tangible reinforcers in the respect that it is a consequence more readily available and appropriate in the regular classroom setting. It would appear that high interest activities in the regular classroom in the form of working on special projects, listening to records, art and craft activities, etc., could be feasibly provided and made contingent upon appropriate academic and social behavior. The use of tangible reinforcers such as candy, trinkets, and toys would appear to be less feasibly implemented in the regular classroom. Special class use of reinforcing events available in the regular class should help to facilitate transition back to full-time regular class placement.

Each child received a work card when he entered the classroom in the morning. During the day the teacher gave "points" for the completion of work assignments and for displaying appropriate social behavior. Each point was worth one minute of free time. The academic task assignments were reinforced on a combined interval-ratio basis. The reinforcement (points) was dispensed at the end of a specified time (interval) but only if the required quantity (ratio) of academic work had been completed.

In the initial stages of bringing the desired behavior under stimulus control, the child was reinforced at frequent intervals for a minimal quantity of academic production. He may, for example, have been reinforced for being ready to learn, starting the assignment, and finishing the task. Gradually these steps were eliminated and reinforcement occurred only at the end of the assigned task. The length of the assigned task was gradually increased up to forty-five minutes.

In similar fashion, the amount of free time was allocated on a gradually decreasing basis. By gradually increasing the intervals between reinforcement and the amount of academic production required and decreasing the amount of reinforcement given, it was possible to establish increasingly high rates of production for minimal amounts of reinforcement. It was interesting to note that the originally low frequency behaviors, academic tasks, often become high frequency behaviors. On numerous occasions the students elected not to take free time, but to read or engage in some other academic task. Such a reversal suggests that the successes obtained in these activities are highly potent reinforcers.

A group reinforcing procedure in which reinforcement is contingent upon the performance of all members of the group was also employed to facilitate the development of productive academic and social behaviors. The group earned points which were exchanged for student selected trips outside the school setting. This procedure is particularly potent since it incorporates positive reinforcers (trips) and aversive consequences (peer disapproval) into the same procedure.

An electric interval timer with a large clock face was utilized in this procedure. The timer operated each day from 11:00 a.m. to 11:30 a.m. A pre-selected interval of time is selected (short initially and gradually increased), and the clock started. The clock remained running as long as all children were engaged in academic task oriented behavior. If day dreaming, talking, or any other behavior incompatible with academic production occurred, the clock was stopped and re-set. When the timer reached zero, the number of points earned (depending on the length of the interval) were entered in bar graph fashion on a large chart visible to all.

The use of positive reinforcement, in the form of individual and group reinforcing climates, to establish and control appropriate academic and social behavior was supplemented by aversive control or punishment. Each academic assignment must have been completed before proceeding to the next. In the event that the student did not finish all his assignments during the class period, he was required to complete the work on his own time at home. The student's admittance to the classroom the following day was contingent upon completion of the assignment.

The use of aversive consequences or punishment to control inappropriate social behaviors was differentially effective depending on the particular procedures employed. Withdrawal of positive reinforcement by removing the child from the classroom contingent upon emission of the inappropriate behavior proved highly effective. Minor disruptions such as talking and wandering around the classroom without permission, throwing objects and swearing resulted in the child being placed in a "time out" room that adjoins the main classroom. Simply stated, "time out" means withdrawing the subject from a positively reinforcing climate. When the child gained control of his behavior and spent a minimum of ten minutes in the "time out" room, he was allowed to return to the main classroom. "Time out from reinforcement" would be expected to be effective only if the reinforcing climate in the experimental classroom is potent enough that the child would rather be there than in the "time out" room.

Fighting, creating a disturbance while in the "time out" room, leaving the classroom without permission, and teacher defiance were consequated by immediate removal from the school setting for at least one full day. As an accompanying consequence, each child who was removed from the school situation was required to complete at home his assignments for those days he was absent. Return to the experimental classroom was made contingent upon successful completion of these assignments.

These two techniques, "time out from reinforcement" and exclusion from the school setting, have proven highly successful in greatly reducing the frequency of inappropriate social behavior.

The amount of productive task oriented behavior engaged in by the students was recorded before, during, and after the operation of the experimental class. Observations of student behavior in the regular classrooms were made prior to the onset of the experimental class to determine the behavioral level maintained by traditional educational procedures and again upon transfer back to the regular classroom to determine what generalization of effect prevailed. A behavior observation form and description of behavioral categories is provided in Appendix D.

Two graduate students independently recorded ten minute samples of each students' behavior on a daily basis during the operation of the experimental classroom. A minimum of six ten minute observations were obtained in the regular classrooms prior and subsequent to the operation of the experimental classroom. Inter-rater reliability checks were performed periodically during the operation of the experimental classroom. The reliabilities were calculated by a percent agreement method where number of agreements are divided by the total number of symbols. Reliabilities ranged from .70 to 1.00 with a median of .86.

A description of the five treatment phases during which behavioral observations were obtained is provided in the following section.

Phase I

(1) Reinforcing Climate - Individual positive reinforcement for good social and academic behavior (points exchanged for free time) on an individual basis.

(2) Academic Consequences - No aversive controls were employed--failure to complete an assignment simply failed to bring reinforcement.

(3) Deviant Social Behavior - Minor disruptions were ignored--major disruptions were consequated by "time out from reinforcement."

Phase II

(1) Reinforcing Climate - Individual basis--same as Phase I.

(2) Academic Consequences - None.

(3) Deviant Social Behavior - Ignored all deviant behavior; no consequences.

Phase III

(1) Reinforcing Climate - Individual--same as Phases I and II.

(2) Academic Consequences - Students required to complete each assignment; all assignments each day must have been completed before the student could enter the class on the following day.

(3) Deviant Social Behavior - Minor disruptions resulted in "time out from reinforcement;" major disruptions resulted in exclusion from school for one full day.

Phase IV

(1) Reinforcing Climate - Individual basis--same as Phase I - III.

(2) Reinforcing Climate - Group Basis--students received bonus points for good academic and social behavior which were exchanged for "special trips."

(3) Academic Consequences - Same as Phase III.

(4) Deviant Social Behavior - Same as Phase III.

Phase V

Regular Classroom - The teachers were introduced to the behavioral control procedures employed in the experimental classroom. An individual program specifying the use of these procedures was provided for each teacher. No steps were taken to insure teacher adherence to the program.

Results

Data presented in Tables I - V show the amount of student task oriented behavior under four different treatment conditions in the experimental classroom

(Phases I - IV) and upon return to the regular classroom (Phase V). Observations were obtained during the first two weeks following the students return to the regular classroom. A six-month follow-up is currently underway.

TABLE I

Percent Task Oriented Behavior for Subject I

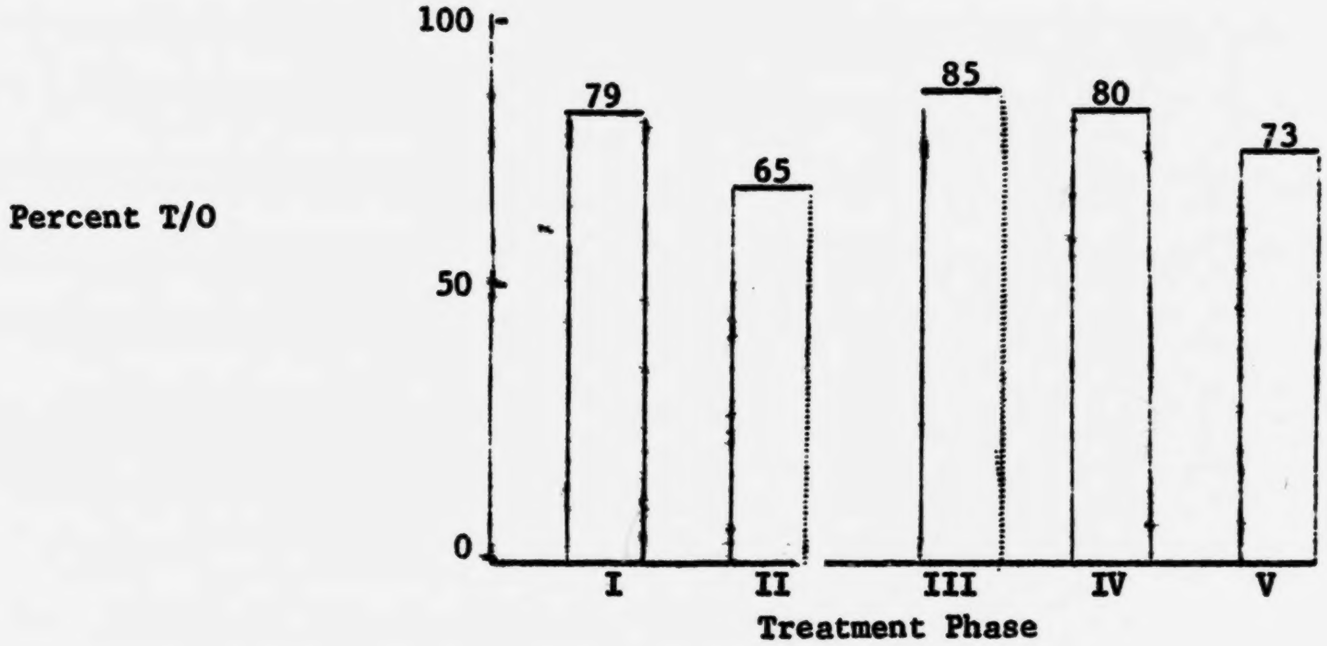


TABLE II

Percent Task Oriented Behavior for Subject II

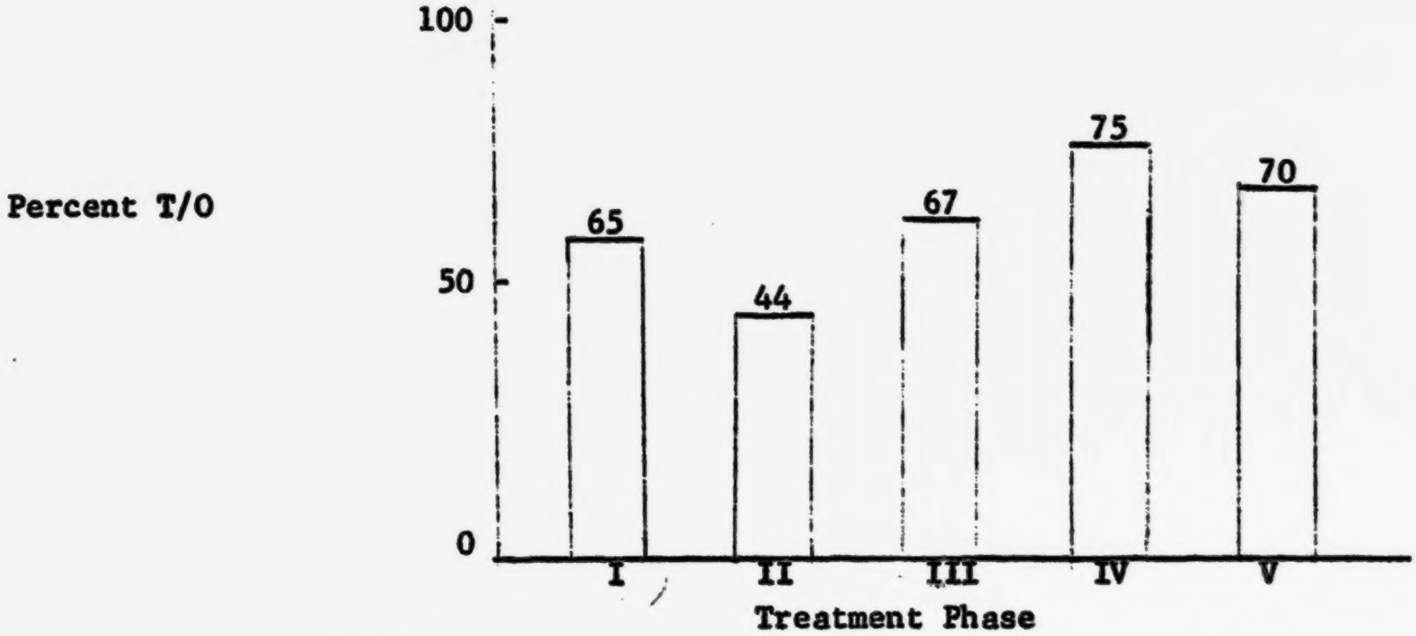


TABLE III

Percent Task Oriented Behavior for Subject III

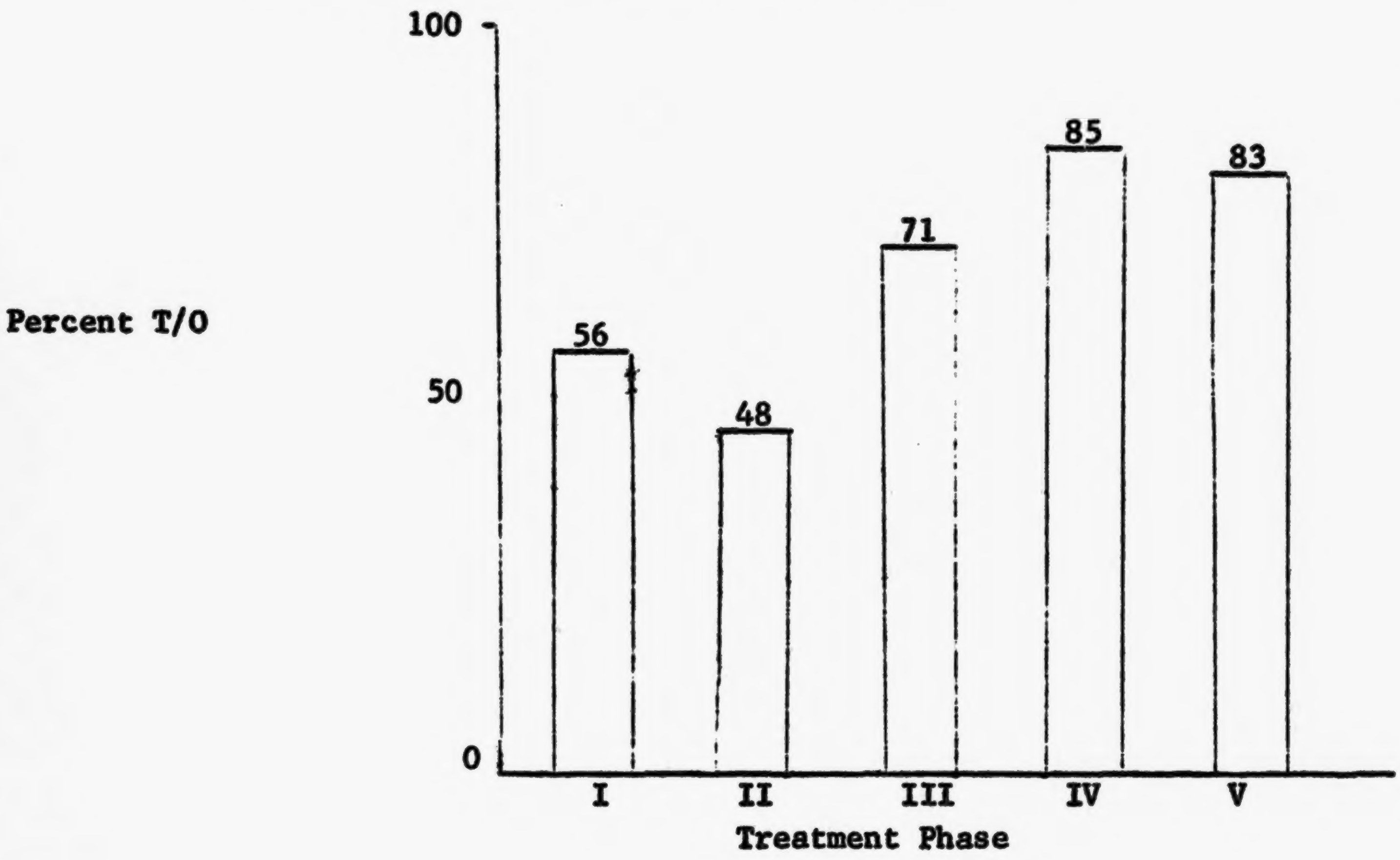


TABLE IV

Percent Task Oriented Behavior for Subject IV

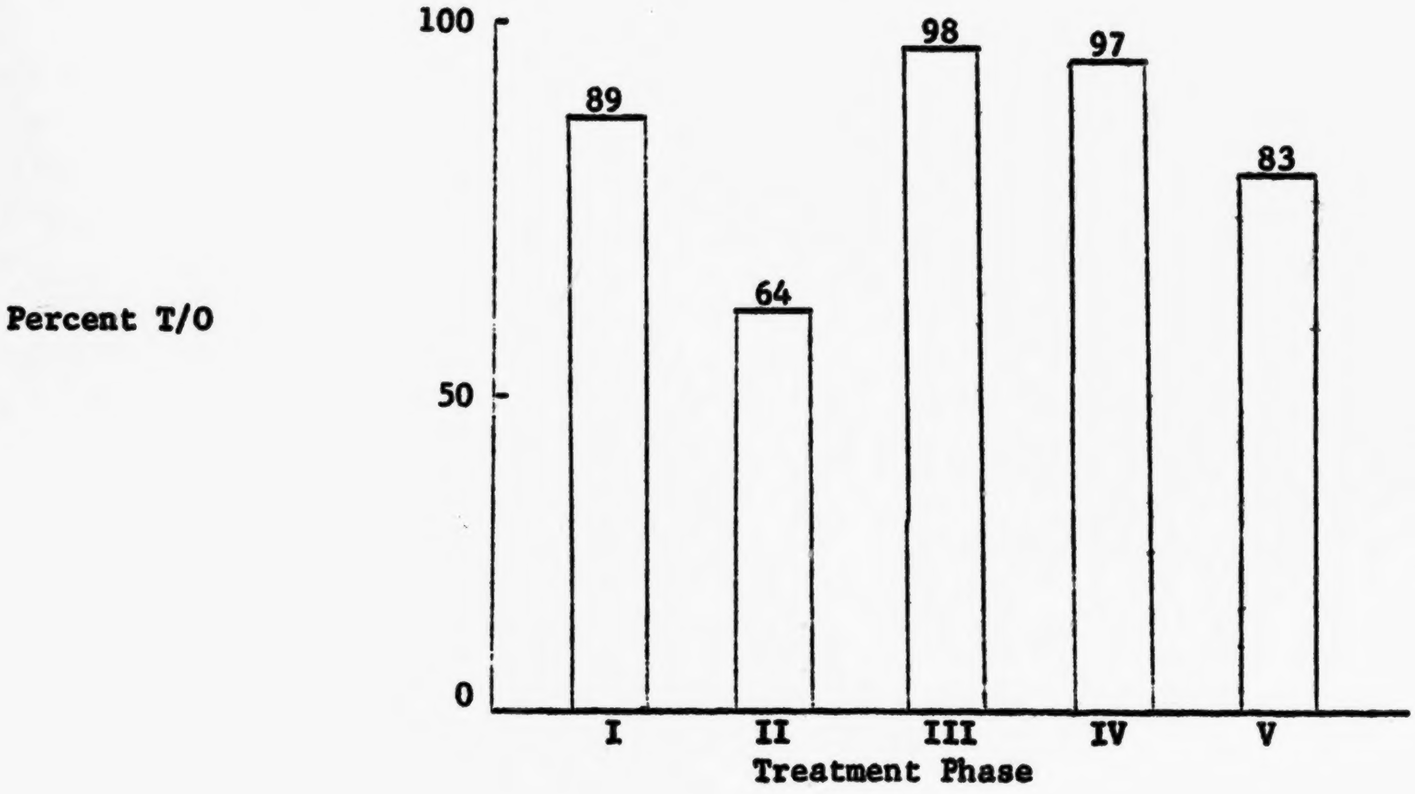
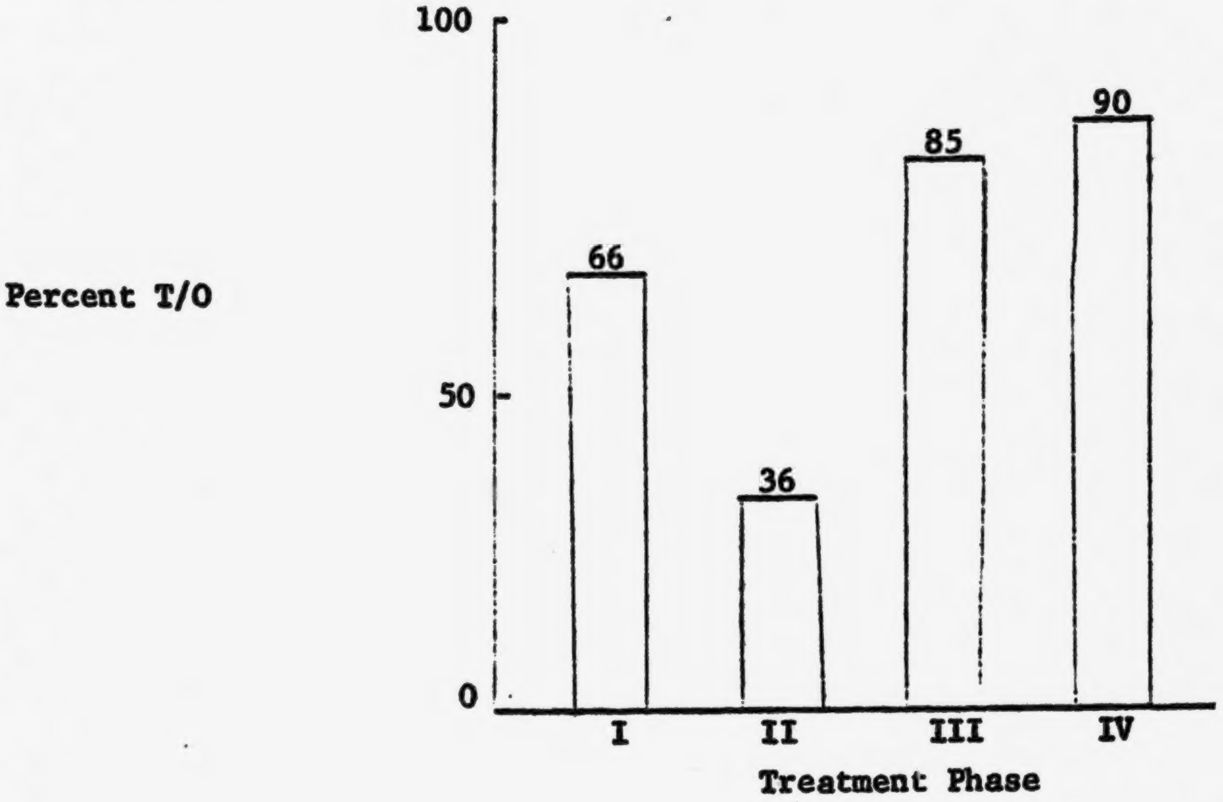


TABLE V

Percent Task Oriented Behavior for Subject V



The available data suggests that the various combinations of treatment procedures employed were differentially effective in producing behavioral change. The positive reinforcing climate present during Phase I was sufficient to maintain a class average of 71% task-oriented behavior during individual study time. During Phase II the consequences for deviant social behavior were removed. Subsequently, inappropriate social behaviors increased and task-oriented behavior decreased to a class average of 51%. It is reaily apparent that a classroom relatively free of behavioral disruption is a necessary prerequisite for efficient prosthetic application of reinforcement procedures to academic behavior. During Phase III the consequences of deviant social behavior were reinstated and expanded to include immediate exclusion from the school setting for major disruptions. Shortly after the initiation of this consequence these behaviors dropped out almost entirely.

In addition, aversive consequences for failure to complete academic assignments within the allotted time were initiated. Recess, free-time, and, in the event that the work was incomplete at the end of the class period, admission to the class the following day, were all made contingent upon completion of the assignment. During this phase the class averaged 81% task-oriented behavior. The 10% increase over Phase I suggests that a combination of positive reinforcement and aversive consequences for academic productivity was more effective than positive reinforcement alone in increasing task-oriented behavior under these conditions. During Phase IV a group reinforcing climate was initiated. It is believed the subsequent gain in total task-oriented time (an increase of 4% over the 81% of Phase III) does not accurately reflect the potential effectiveness of this procedure since the student's behavior was already under a high degree of control. This procedure was observed to demonstrate a very high degree of control over student behavior during the daily 30 minute sessions. The efficiency of this technique will undoubtedly be more clearly demonstrated when it is employed in the initial stages of gaining behavioral control.

Method - Group II

The same criteria were employed in the selection of the second group of students. Each student was enrolled in the fourth, fifth, or sixth grade, average or above in intellectual ability, one or more years retarded in a basic skills area and displayed a high frequency of acting-out behavior. In order to comply with all selection criteria, it was necessary to accept students from distant elementary schools in the district as a sufficient number of children meeting the criteria were not available in the two schools within walking distance of the school where the experimental classroom was located. It was therefore necessary to bus these children between their homes and the experimental classroom. This busing arrangement required that the program be shifted from a half day regular and half day experimental class placement to a full day placement in the experimental class.

There were several other important modifications in the program. The "back up" reinforcers for the points that students received individually for good student behavior and academic production were changed from earned time to tangible objects such as models and games. Following the procedures employed with Group I students, the points earned collectively by the group were exchanged for group "earned time" activities of high interest, such as swimming, bowling, and playing ping-pong.

The various response-reinforcement contingencies employed during the four treatment phases with Group I were modified and incorporated into one set of procedures which were used exclusively with Group II. A description of these procedures follows:

1. Reinforcing Climate - Individual-Positive reinforcement for good social and academic behavior (points exchanged for models).
2. Reinforcing Climate - Group Basis-Students received points for appropriate academic and social behavior which were exchanged for "special trips" and activities.
3. Academic Consequences -- Students were required to complete each assignment; all daily assignments must have been completed before the student could return to the class on the following day.
4. Deviant Social Behavior -- Minor disruptions resulted in "time out from reinforcement;" major disruptions resulted in exclusion from school for one full day.

The specific methodologies involved in implementing the various response-reinforcement contingencies were identical to those described for Group I. The duration of the program was seven weeks. The data provided in Tables 6 - 10 indicate the effect of the treatment program on student task oriented behavior.

Figure 1

Percent Task Oriented Behavior for Student I

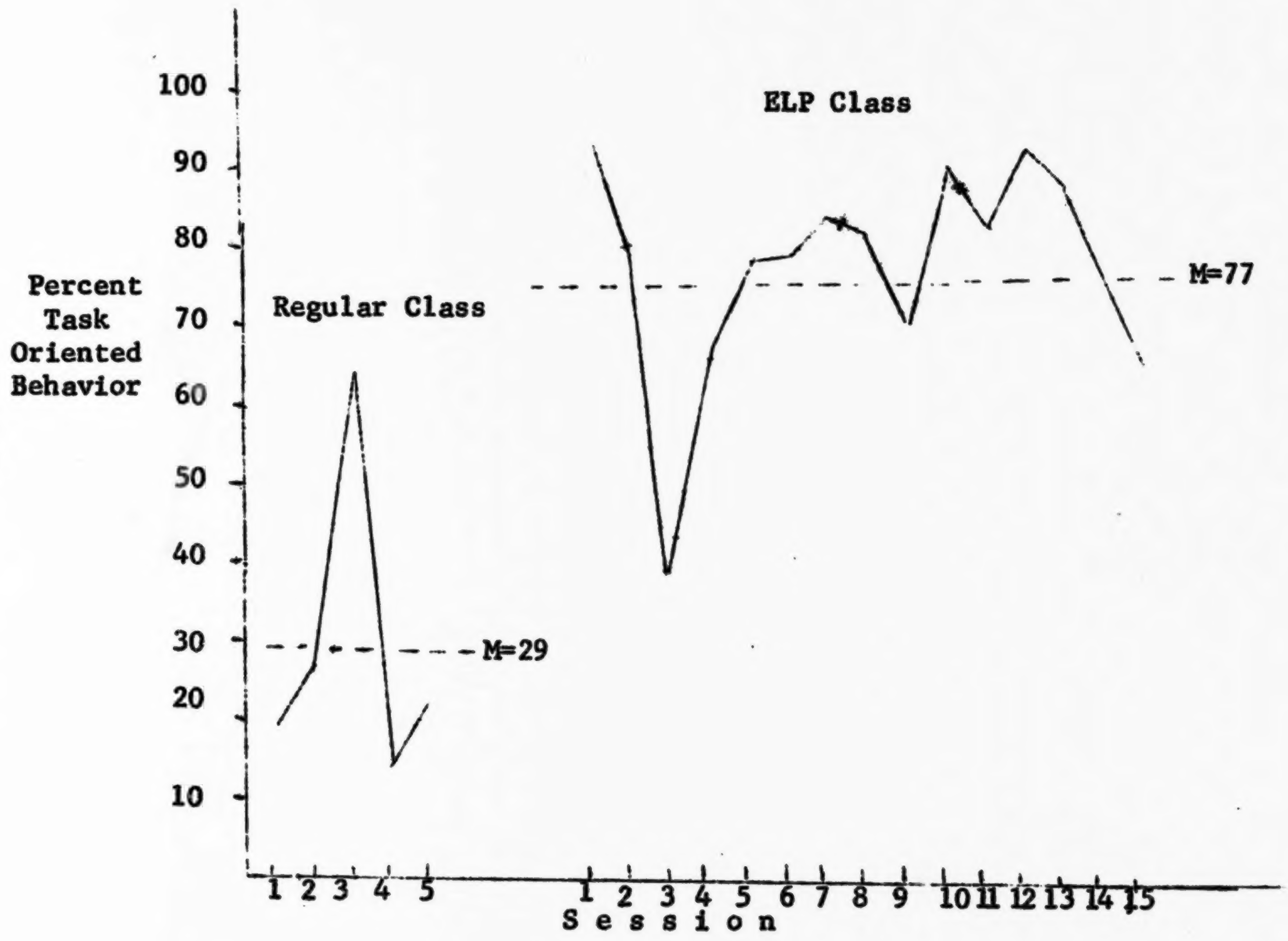


Figure 2

Percent Task Oriented Behavior for Student 2

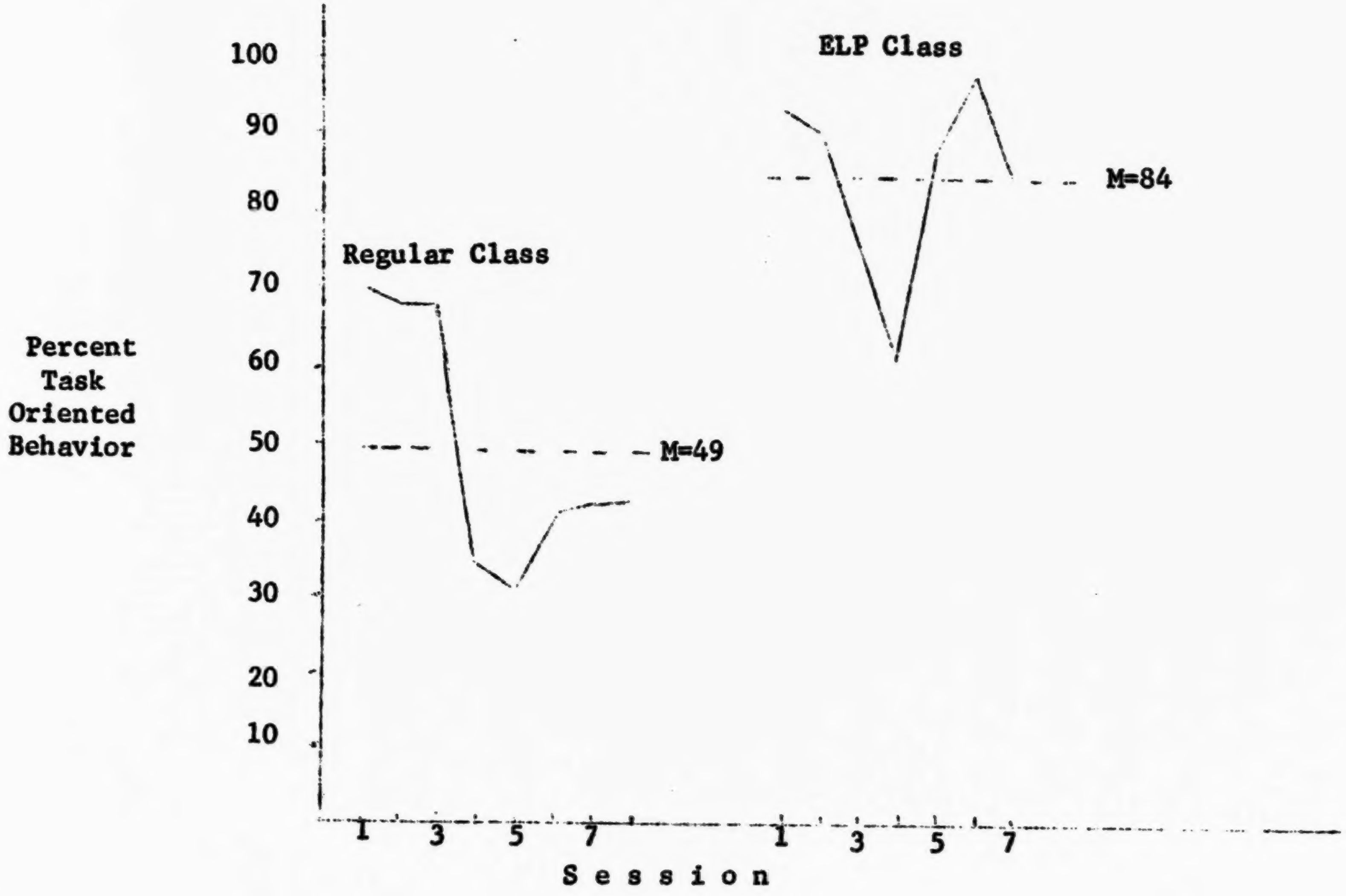


Figure 3

Percent Task Oriented Behavior for Student 3

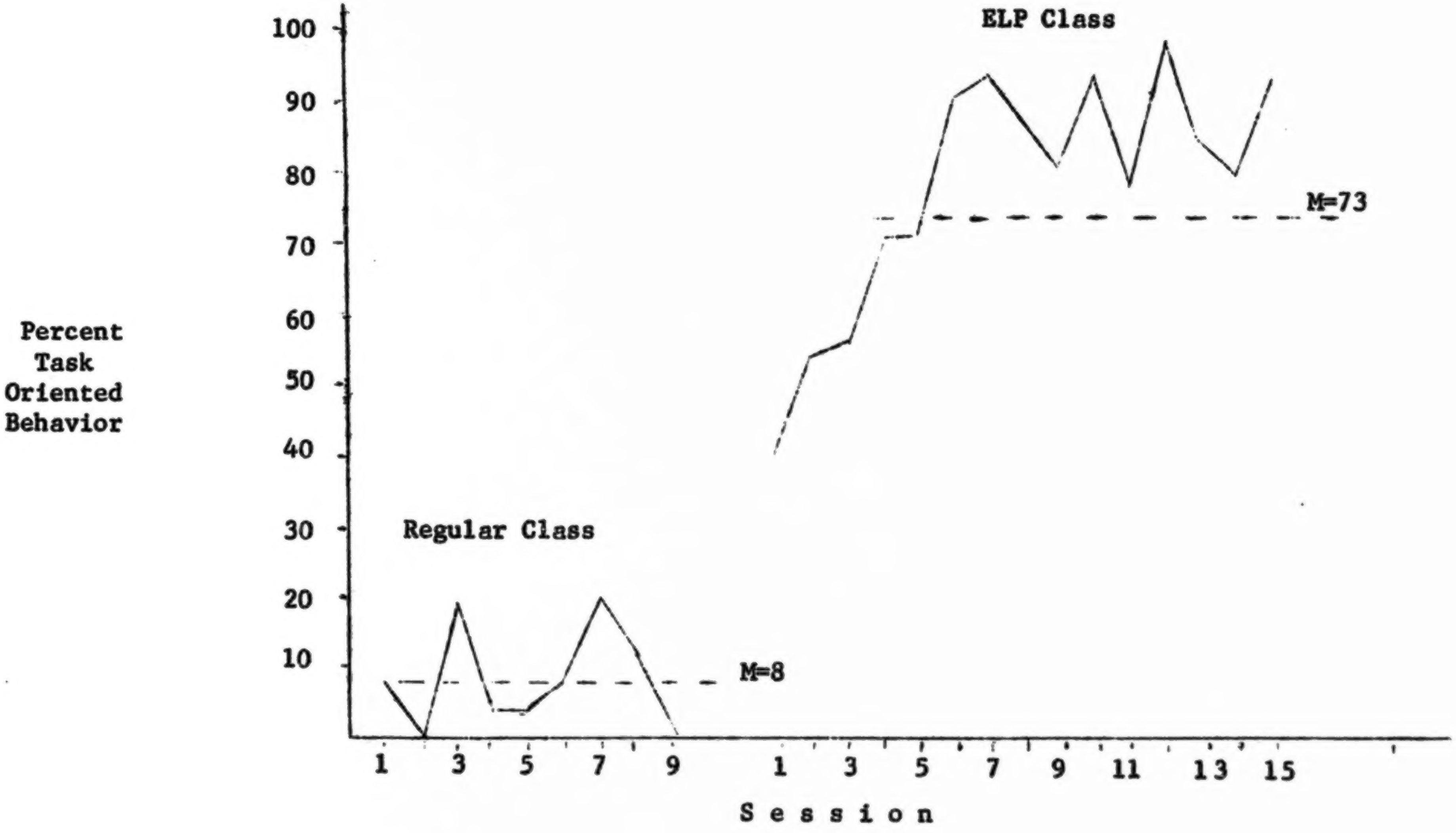


Figure 4

Percent Task Oriented Behavior for Student 4

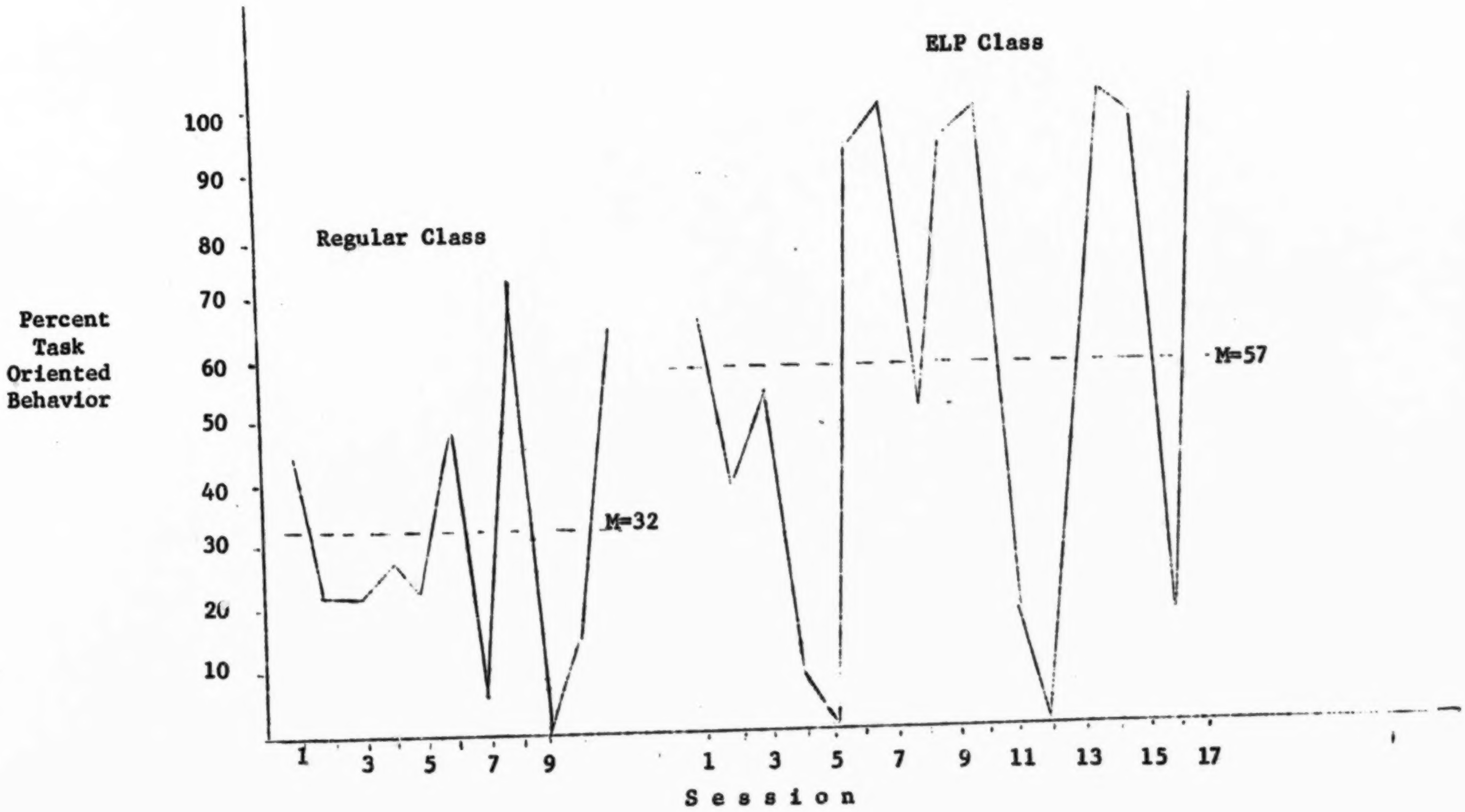


Figure 5

Percent Task Oriented Behavior for Student 5

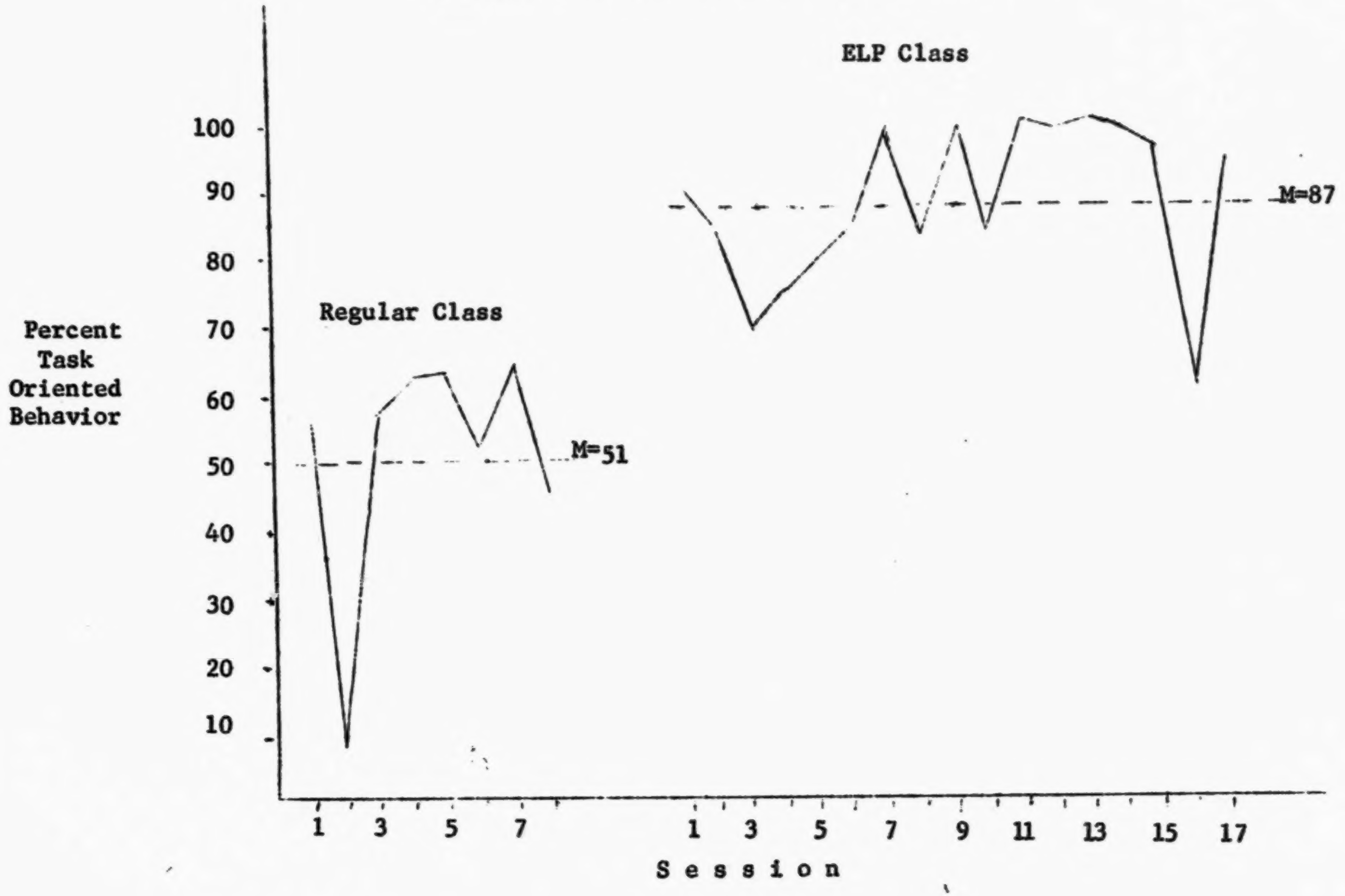
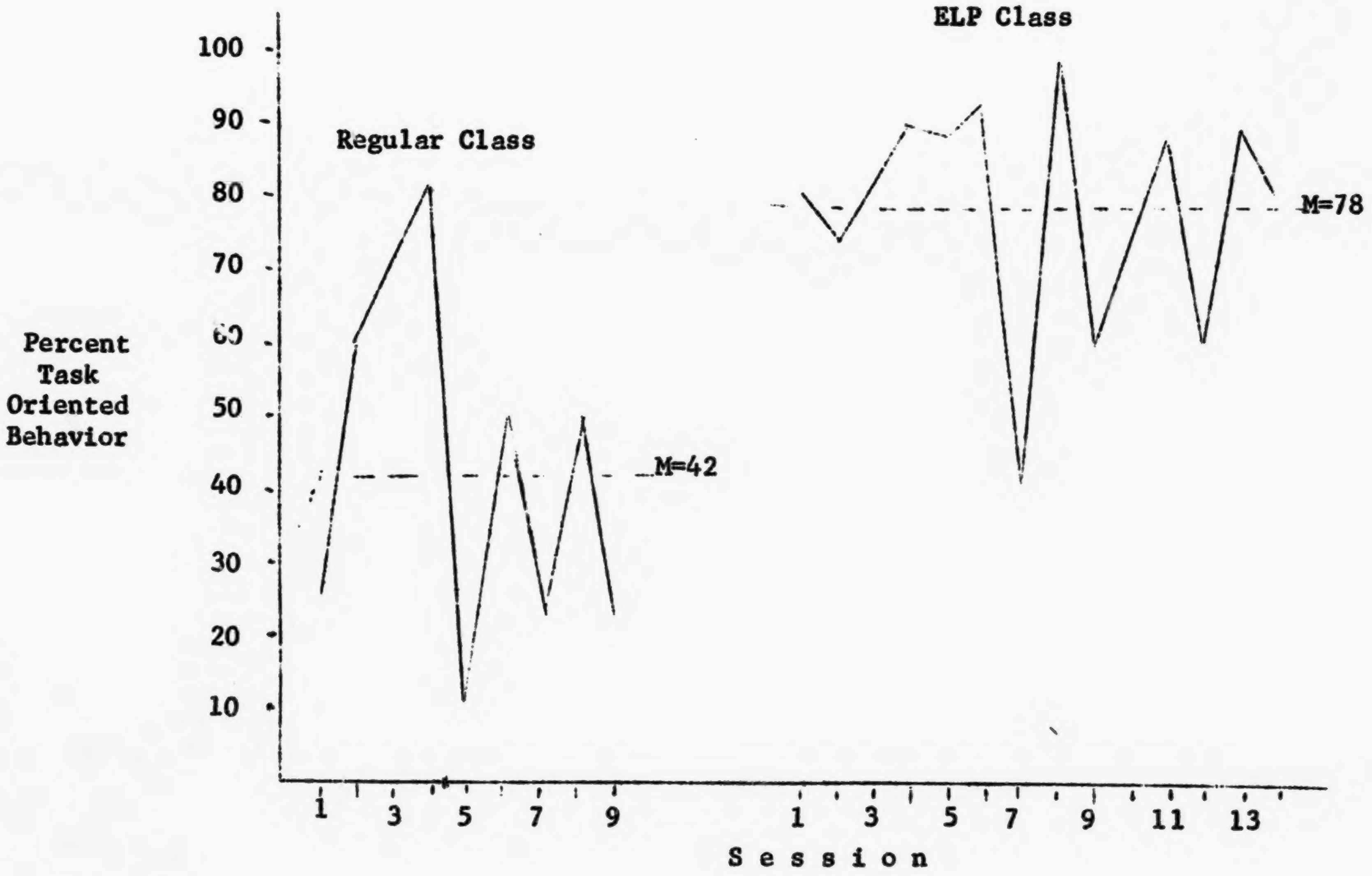


Figure 6

Percent Task Oriented Behavior for Student 6



Discussion

The data provided in Tables I - V suggest that the educational procedures employed in the special class were differentially effective in producing a marked increase in student task oriented behavior. In addition, preliminary follow-up observations revealed that a high rate of productive academic behavior was maintained upon the student's return to the regular classroom.

The development and implementation of these treatment procedures was a preliminary exploratory effort and as such did not involve a high degree of procedural experimental control. As a result, the data do not provide a basis for a valid experimental appraisal of the treatment variables employed.

As can be seen in Figures 1 - 6, the amount of task oriented behavior increased appreciably for all students during the operation of the experimental classroom. The greatest increase in productive academic behavior was shown by Student 3 whose task oriented behavior increased from 8% during regular classroom conditions to an average of 73% under experimental classroom conditions. The smallest gain was shown by Student 4 with increases from an average of 32% under regular classroom conditions to an average of 57% under experimental classroom conditions. Inspection of Figure 4 reveals that the academic behavior of Student 4 was highly erratic. This student was typically either completely involved in the academic task or completely uninvolved or non-task oriented, indicating that the response reinforcement contingencies in operation had established only tenuous control over Student 4's academic behavior.

The data presented in Figures 1 - 6 indicate that task oriented behavior increased for all students under the conditions operating during their seven week enrollment in the experimental classroom. It should be noted, however, that these findings (as with Group I) do not represent a valid appraisal of the effect of the treatment variables on task oriented behavior. It is possible, for example, that setting and teacher variables specific to the experimental classroom had an equal or greater effect on task oriented behavior than did the token reinforcement system and other contingency relations. A paradigm that involves (1) the establishment of a stable base rate of behavior, (2) the manipulation of an experimental variable resulting in a change in behavioral rate, and (3) the withdrawl, alteration, or reversal of contingencies resulting in a return to baseline conditions is a necessary prerequisite to valid appraisal of treatment variables. The establishment of functional relationships between treatment variables and behavioral variables is based upon the use of such experimental controls. Promising data derived from designs lacking in such controls must be regarded as preliminary and non-conclusive. Future project efforts will be directed toward an experimental evaluation of the treatment procedures employed with Groups I and II.

Current educational practice appears to reflect a belief that behavior change in one setting will transfer or generalize to other settings. It is not uncommon practice, for example, for educators to deal with a child's inappropriate academic and social behaviors in clinics and special classes and then assume generalization to the regular classroom. These settings differ in many important respects. The response-reinforcement contingency relations and schedules of reinforcement initially responsible for shaping and maintaining the inappropriate behavior are absent from the treatment setting. It is not surprising, therefore, that "generalization" of effect is often limited and, when present, difficult to account for.

A critical aspect of the development of educational procedures for these children involves the identification and control of the variables responsible for behavioral transfer, i.e., those conditions specific to the individual and to the regular class setting that serve to maintain the desired behavior. In order to insure a transfer of behavior from the special to regular class setting, therapeutic efforts would do well to focus on (1) attempts to re-program the regular class environment and (2) shifting behavioral control from artificial to natural reinforcers.

Attempts to re-program the regular classroom environment focusing upon the alteration of those response-reinforcement contingencies and reinforcement schedules identified as being related to the student's inappropriate behavior should increase the degree of behavioral transfer. Another factor that appears to contribute substantially to behavioral transfer is the degree to which the behavior of the student has come under the control of "natural" as opposed to "artificial" reinforcers. Natural reinforcers are those that possess a logical link to the behavior that they follow. The skillful use of eating utensils is reinforced by the accurate maneuvering of food into the mouth. Food in this instance is a natural reinforcer. If food is used to reinforce a child for sitting still, it is not being used as a natural reinforcer.

This distinction between "natural" and "artificial" reinforcers has important implications for maintaining behavior in the regular school environment.

Natural reinforcers are generally more available and permanent in their effect on maintaining behavior than artificial or contrived reinforcers. There are some students, however, whose poor academic work and aversive social behavior precludes or severely limits the availability of the "natural" reinforcers normally present and available. In such instances, "artificial" reinforcers such as those employed in a token economy system serve a highly useful function. Desired academic and social behavior typically comes under rapid control of these "artificial" reinforcers. Once the appropriate behavior becomes a part of the child's repertoire, previously unavailable natural reinforcers become available and assume the behavioral control function. A child who experiences great difficulty in reading, for example, typically reads only infrequently. The avenues of reinforcement which result from successful reading efforts (learning new facts, task completion, social approval, etc.) are unavailable to him. The use of tokens or other artificial reinforcers to gain initial control over the reading behavior increases the availability of more appropriate reinforcers intrinsic to the reading process.

Similar efforts aimed at shifting the control of other academic and social behaviors from artificial to natural reinforcers should further increase behavioral transfer to the regular class setting.

SUMMARY

The research project is divided into two sections: (1) The first section focuses upon developing assessment instruments for the identification of disturbed children. (2) The second section is concerned with developing a treatment model that will be effective in modifying the behavior of disturbed children in the educational setting.

A behavior checklist, a behavior rating scale, and a behavioral observation form have been constructed for the purpose of fulfilling objective one. Procedures on validating and estimating the reliability of the checklist have been completed. The split-half reliability estimate is .98. The scale discriminates between disturbed and non-disturbed children at the .001 level of confidence. Scores on the checklist (N = 534) correlate .68 with a criterion of behavior disturbance. Preliminary data on the rating scale indicates that the scale reflects treatment differences which are known to exist - p = .01. The average inter-rater reliability for three judges on the behavior of six subjects was .935. Agreement measures between independent observers using the behavioral observation form are .90 and above.

The treatment model, based upon learning theory principles, has produced measurable behavior change in disturbed fourth, fifth, and sixth grade male subjects. The researchers are not in a position at this time to indicate which treatment variables are producing a given amount of behavior change. The ensuing year will be spent in determining the weight which each specifiable treatment variable exerts upon the dependent variable of behavior change. Changes recorded to this writing indicate reduced frequencies of deviant behavior and increased proportions of time spent engaged in task oriented behavior.

BIBLIOGRAPHY

Beilin, H. "Teachers and Clinicians Attitudes Toward the Behavior Problems of Children"--A Reappraisal, Child Development, Vol. 30, 9-25.

Bower, Eli. "A Process for Early Identification of Emotionally Disturbed Children," Bulletin of the California State Department of Education, Vol. 27 (1958), 1-111.

Bruekner, Leo J., and Bond, Guy L. The Diagnosis and Treatment of Learning Difficulties. New York: Appleton-Century Crofts, Inc., 1955.

Buros. Mental Measurements Yearbook. Fifth Edition, 1959.

Committee on Child Psychiatry. Psychopathological Disorders in Childhood: Theoretical Considerations and a Proposed Classification. New York: Group for the Advancement of Psychiatry, 1966.

Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Row, 1960.

Dreger, R., et. al. "Behavior Classification Project," Journal of Consulting Psych., Vol. 28 (1964), 1-12.

Garrett, Henry E. Statistics in Psychology and Education. New York: David McKay Company, Inc., 1962.

Graves, Palmer A. "Facilitating the Results of Therapy," Elementary School Journal, Vol. 62 (December, 1957), 166-167.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Co., Inc. 1954.

Harper, B. P., and Dunlap, J. W. "Derivation and Application of a Unit Scoring System for the Strong Vocational Interest Blank for Women," Psychometrika, Vol. 7 (1942), 289-295.

Havighurst, Robert J. "Dealing with Problem Youth," The Nation's Schools, Vol. 61 (May, 1958), 43-45.

Hunter, E. C. "Changes in Teachers' Attitudes Toward Children's Behavior Over the Last Thirty Years," Mental Hygiene, Vol. 41 (1957), 3-11.

Jordon, Thomas E. The Exceptional Child. Charles E. Merrill, 1962, pp. 250-253.

Kelley, T. L. "The Selection of Upper and Lower Groups for the Validation of Test Items," Journal of Educational Psychology, Vol. 30 (1939), 17-24.

Kirk, Samuel A. Educating Exceptional Children. Houghton-Mifflin, 1962, pp. 350-353.

Kvaraceus, W. C. "Forecasting Juvenile Delinquency; Supplement to the Manual of Directions for the KD Proneness Scale and Checklist," Journal of Education, Vol. 138 (April, 1956), 1-43.

Laycock, Samuel. Mental Hygiene in the School. Toronto: Copp Clark Publishing Co., ltd., 1960.

Levitt, Eugene, E. Clinical Research Design and Analysis in the Behavioral Sciences. Springfield: Charles C. Thomas, Publisher, 1961.

Likert, R. "A Technique for the Measurement of Attitude Scales," Arch. Psych., New York, 1932, No. 140.

Lindquist, E. F. Educational Measurement. Washington, D. C.: George Banta Publishing Co., 1950.

Maes, Wayne R. "The Identification of Emotionally Disturbed School Children," Vol. 32 (May, 1966), 607-613.

Newman, R. G. "Measuring the Learning and Adjustment of Six Hyperactive Males," American Journal of Ortho-Psychiatry, Vol. 29, 637-643.

Phillips, E. Lakin. "Logical Analysis of Childhood Behavior Problems and Their Treatment," Psychological Reports, Vol. 9, (1961), 705-712.

Phillips, A. J. "Further Evidence Regarding Weighted Versus Unweighted Scoring of Examinations," Educational Psychological Measurement, Vol. 3 (1943), 151-155.

Sarason, Seymour R., et. al. Anxiety in Elementary School Children. New York: John Wiley and Sons, Inc., 1960.

Schrupp, M. H., and Gjerde, C. M. "Teacher Growth in Attitudes Toward Behavior Problems of Children," Journal of Educational Psychology, Vol. 44 (1953), 203-214.

Stouffer, George A. W., Jr. "Behavior Problems of Children as Viewed by Teachers and Mental Hygienists: A Study of Present Attitudes as Compared with Those Reported by E. K. Wickman," Mental Hygiene, Vol. 36 (1952), 271-285.

Strang, Ruth, "Mental Health," Chester W. Harris, Editor, Encyclopedia of Educational Research, 3rd Ed., New York: MacMillan Co., 1960, 823-835.

Strong, Edward K. Jr. "Weighted Versus Unit Scales," J. Educational Psychology, Vol. 36 (1945), 193-216.

Thorndike, Robert L. and Hagen, Elizabeth. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1961.

Trippe, Matthew, "Conceptual Problems in Research on Educational Provisions for Disturbed Children," Exceptional Children (April, 1961).

Ullman, Charles A. "Identification of Maladjusted School Children," Public Health Monograph No. 7, Washington, D. C.: Federal Security Agency, 1952.

White and Harris, Myron W. The School Psychologist. New York: Harper and Bros., 1961.

Wickman, E. K. Children's Behavior and Teacher's Attitudes. New York: The Commonwealth Fund, 1928.

Wilks, S. S. "Weighted Systems for Linear Functions of Correlated Variables When There is No Dependent Variable," Psychometrika, 111 (1938), 23-40.

APPENDIX A

Rating Form and Accompanying Instructions for Behavior Science Panel

A list of behavior problems has been tabulated on the attached sheets. The list was collected by way of abstracting operational statements about observable behaviors from teachers' descriptions of behavior problems of children.

You are requested to rate each of these behavioral items according to their importance or influence in handicapping a given child's adjustment. In other words, how undesirable is it for any child to manifest the behavior described in this list of problems?

Your rating will be made by marking on a scale of seriousness provided for this purpose. A line has been drawn to the right of each behavioral item. Each line has four divisional points coinciding with captions at the top of the page to denote the degree of difficulty (or seriousness) caused by such behavior in any child.

Directions:

--First read the list of behavioral items on the two sheets.

--Then answer the questions of the degree of seriousness and undesirability of each item when manifested by any child by making a vertical stroke at any point on the line according to the captions at the top of the page.

--You may make your rating at any point on the line. You do not need to make it directly on any divisional point. If you think your rating falls somewhere between two divisional points, make your mark at the appropriate point where you think it ought to go. This will permit you to distinguish finely in your ratings between the different behavior problems. The following example will illustrate this:

How serious (or undesirable) is this behavior in any child?

No Importance Slight Importance Medium Importance Great Importance

Stealing . . . . ____

--Avoid rating how frequently the particular behavior occurs in children. Rate only how undesirable it is for any child when it does occur.

--Do not ponder an excessive amount of time over any one item.

(Wickman, 1928, p. 206)

BEHAVIOR CHECKLIST

Pupil Name ______________________ LAST ______________ FIRST ______________ Middle Initial

School ______________________ Pupil age ________

Pupil sex		BOY ☐	GIRL ☐								
Grade in school									☐ 4	☐ 5	☐ 6
Sex of Rater		☐ M	☐ F								
SPECIAL CODES		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9
		☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9

Please read each item carefully and respond by marking "yes" or "no" as it applies to the child. If you have observed a particular behavior enough to know that it is part of the child's behavioral response pattern and not just a chance occurrence, answer the item by marking in the "yes" column. If you have not observed the behavior in the child, mark in the "no" column. Mark either "yes" or "no" for each item. Do not omit any.

		YES	NO
1	Complains about others unfairness and/or discrimination toward him.	☐	☐
2	Is listless and continually tired.	☐	☐
3	Does not conform to limits on his own without control from others	☐	☐
4	Becomes hysterical, upset, or angry when things do not go his way.	☐	☐
5	Comments that no one understands him.	☐	☐
6	Perfectionistic: Meticulous about having everything exactly right.	☐	☐
7	Will destroy or take apart something he has made rather than show it or ask to have it displayed.	☐	☐
8	Other children act as if he were taboo or tainted.	☐	☐
9	Has difficulty concentrating for any length of time.	☐	☐
10	Is overactive, restless, and/or continually shifting body position.	☐	☐
11	Apologizes repeatedly for himself and/or his behavior.	☐	☐
12	Distorts the truth by making statements contrary to fact.	☐	☐
13	Underachieving: Performs below his demonstrated ability level.	☐	☐
14	Disturbs other children: teasing, provoking fights, interrupting others.	☐	☐
15	Tries to avoid calling attention to himself.	☐	☐
16	Makes distrustful or suspicious remarks about actions of others toward him.	☐	☐
17	Reacts to stressful situations or changes in routine with: general body aches, head or stomach aches, nausea.	☐	☐
18	Argues and must have the last word in verbal exchanges.	☐	☐
19	Approaches new tasks and situations with an "I can't do it" response.	☐	☐
20	Has nervous tics: muscle-twitching, eye-blinking, nail-biting, hand-wringing.	☐	☐
21	Habitually rejects the school experience through actions or comments.	☐	☐
22	Has enuresis.	☐	☐
23	Utters nonsense syllables and/or babbles to himself.	☐	☐
24	Continually seeks attention.	☐	☐
25	Comments that nobody likes him.	☐	☐
26	Repeats one idea, thought, or activity over and over.	☐	☐
27	Has temper tantrums.	☐	☐
28	Refers to himself as dumb, stupid, or incapable.	☐	☐
29	Does not engage in group activities.	☐	☐
30	When teased or irritated by other children, takes out his frustrations on another inappropriate person or thing.	☐	☐
31	Has rapid mood shifts: depressed one moment, manic the next.	☐	☐
32	Does not obey until threatened with punishmenr.	☐	☐
33	Complains of nightmares, bad dreams.	☐	☐
34	Expresses concern about being lonely, unhappy.	☐	☐
35	Openly strikes back with angry behavior to teasing of other children.	☐	☐
36	Expresses concern about something terrible or horrible happening to him.	☐	☐
37	Has no friends.	☐	☐
38	Must have approval for tasks attempted or completed.	☐	☐
39	Displays physical aggression toward objects or persons.	☐	☐
40	Is hypercritical of himself.	☐	☐
41	Does not complete tasks attempted.	☐	☐
42	Doesn't protest when others hurt, tease or criticize him.	☐	☐
43	Shuns or avoids heterosexual activities.	☐	☐
44	Steals things from other children.	☐	☐
45	Does not initiate relationships with other children.	☐	☐
46	Reacts with defiance to instructions or commands.	☐	☐
47	Weeps or cries without provocation.	☐	☐
48	Stutters, stammers or blocks on saying words.	☐	☐
49	Easily distracted away from the task at hand by ordinary classroom stimuli, i.e. minor movements of others, noises, etc.	☐	☐
50	Frequently stares blankly into space and is unaware of his surroundings when doing so.	☐	☐

BEHAVIOR RATING SCALE

Demographic Information:

Name of Pupil__________________________ Date of Birth __________

School__________________________________ Grade __________________

Sex of Rater___________________________ Sex of Pupil____________

Name of Rater__________________________ Date____________________

Instructions to the rater:

1. This scale is designed for the purpose of identifying behaviorally disturbed children. Items in the scale represent OVERT BEHAVIORS WHICH CAN BE VERIFIED BY OBSERVATION. Thus, if you have not observed a particular behavioral item in the classroom, you would indicate in the scoring section that the behavior had never occurred.

2. In the first part of the scale, three rating judgments are required for each behavioral item: (a) rate of occurrence (b) rater response (c) rater reaction. One judgment is required under (a) rate of occurrence; one judgment is required under (b) rater response; and one judgment is required under (c) rater reaction. Thus, there would not be more than three rating judgments per item.

   Rate of occurrence is designed to secure information on the frequency with which a particular behavior occurs within the classroom setting. For example, if a behavior occurs one or more times in a week, you would place a check (✓) in box 3 under rate of occurrence.

   Rater response determines how you respond to different behaviors as they occur within the classroom setting. For example, you may respond to a behavior such as not paying attention with a warning glance. On the other hand, you may respond to fighting by temporarily removing the child from the classroom setting. Under rater response, you are asked to indicate how you respond to different behaviors as they occur within the classroom by indicating which of the techniques under rater response you typically use in coping with the behaviors listed in this scale. It is recognized that you use different techniques with the same behavior, depending upon the situation; but you are asked to indicate which technique you usually or typically use in coping with the behavior in question.

   Rater reaction indicates how you, as the rater, react to the differential behaviors exhibited by school children. For example, if a child constantly defies you, are you not disturbed by this behavior, or does it disturb you to a very great extent?

3. Rate the items in the first part of the scale as follows: If you have observed a particular behavior in the classroom, place a check (✓) in the appropriate boxes after that item. If you have not observed a given behavior in a child, place a check in the (0) box under rate of occurrence and leave the other two sections (Rater response and Rater reaction) blank for that item. In the second part of the scale, simply indicate the frequency with which behaviors occur that you have observed. Read all items carefully and respond to every item in the scale.

4. Indicate your judgments in each of the three scoring areas according to the following criteria.

Section A: Rate of Occurrence

(0) The behavior has never occurred.
(1) The behavior occurs at least once every two months.
(2) The behavior occurs at least once a month.
(3) The behavior occurs at least once a week.
(4) The behavior occurs at least once a day.
(5) The behavior occurs at a constant or near constant rate.

Section B: Rater Response: When this particular behavior occurs, do you
(1) Ignore the behavior?
(2) Give the child a warning glance?
(3) Interact verbally or physically with the child?
(4) Temporarily remove the child from the classroom setting?
(5) Refer the child to an outside source, i.e., counselor, psychologist, or separate referral agency?

Section C: Rater Reaction

(1) The behavior does not disturb you.
(2) The behavior disturbs you to a slight extent.
(3) The behavior disturbs you to a moderate extent.
(4) The behavior disturbs you to a great extent.
(5) The behavior disturbs you to a very great extent.

5. Enter appropriate criticisms about the design, item wording, format, and/or directions of this instrument.

6. Sample item:

	Section A Rate of Occurrence						Section B Rater Response					Section C Rater Reaction				
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1. Shouts back when corrected in class.			✓				✓							✓		

This behavior is rated as: occurring at least once a month; the rater ignores the behavior; the behavior is moderately disturbing to the rater.

PART ONE

	Rate of Occurrence						Rater Response					Rater Reaction				
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1. Does not obey commands or directives.																
2. Willingly accepts challenges and gets into fights.																
3. Terminates an irritating or inappropriate behavior if verbally reprimanded, only to resume the behavior when he is not being observed.																
4. Goes through other children's possessions without authorization.																
5. Creates a disturbance during class activities in which he is not interested or skilled.																
6. Responds to teasing with physical aggression.																
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
7. Pouts.																
8. Provokes other children in the classroom by disturbing, teasing or shoving them.																
9. Does not play in games with other children.																
10. When angry, slams books on the desk, slams doors, kicks chairs, etc.																
11. Does not attend to a given task when asked to do so.																
12. Uses profane language in the classroom.																
13. Makes verbal statements such as: You can't make me do this!																

	Rate of Occurrence						Rater Response					Rater Reaction				
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
14. Initiates fights with other children.																
15. Refuses to do any school work for a period of time.																
16. Comments that he hates his teacher.																
17. Attempts to yell the teacher down in front of the class.																
18. If the teacher insists that he do school work when he has refused, throws a temper tantrum, cries, screams, etc.																
19. Argues and demands the last word.																
20. Provokes fights on the playground, reports the fight, then denies having initiated the fight.																
21. Leaves the classroom without permission.																
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
22. Will destroy or take apart something he has made rather than show it or ask to have it displayed.																
23. Refuses to perform or speak before the group when requested.																
24. Threatens other children with physical violence.																
25. Screams, bangs objects when denied something.																
26. Attacks other children with potentially dangerous objects: knives, pencils, sharp objects, etc.																
27. Proceeds to do things before instructions are finished.																
28. When angry, will destroy his own possessions: books, models, pencils, paper, etc.																

	Rate of Occurrence						Rater Response					Rater Reaction				
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
29. Does not follow rules of games, class activities.																
30. Refuses to recognize the fact when he is proven mistaken or wrong.																
31. Does not mind or obey until physically punished.																
32. Threatens to call in his parents to extricate himself from a hostile interaction with the teacher.																
33. Protests about changes in his routine.																
34. Makes loud verbal outburst without raising his hand and securing permission to speak.																
35. Requires control from others before conforming to limits.																
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
36. Cries when things do not go his way.																
37. Ignores warnings and reprimands.																
38. Steals things from other children.																
39. Encourages destructive activity or disobedience in others.																
40. Destroys or defaces property other than his own.																
41. Comments that he hates school.																
42. Forces the teacher to give him her attention.																
43. Displays violent temper tantrums.																
44. Refuses to recite aloud in class.																
45. Engages in fights on the playground.																
46. Does not express himself orally.																

	Rate of Occurrence						Rater Response					Rater Reaction				
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
47. Does not enter into relationships with other children																
48. Strikes another child and then leaves, not staying to carry on with the other child.																
49. Makes lewd gestures.																
50. Interrupts other children while they are working.																
51. Shouts back when corrected in class.																
52. Pesters other children.																
53. Manipulates other children in order to get them to do what he wishes.																
54. Imitates the behavior of his classmates in a mechanical fashion.																
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
55. Does not follow directions given by the teacher but will follow directions contained in a textbook or assignment.																
56. Asks to be excused from activities in which he is required to participate.																
57. Tattles on other children.																
58. When mistreated by other children, takes out his frustrations on another inappropriate person or thing.																
59. Makes contrary to fact statements.																
60. Corrects other children.																
61. Threatens to kill others.																
62. Picks on smaller or weaker children.																

	Rate of Occurrence						Rater Response					Rater Reaction				
	0	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
63. Teases other children																
64. Tries to settle disagreements aggressively, e.g., by bullying or yelling.																

PART TWO

	Rate of Occurrence					
	0	1	2	3	4	5
65. Starts many activities, but does not finish them.						
66. Complains of headaches, cramps, general body aches.						
67. Uses his hands in a clumsy fashion.						
68. Does not respond to verbal inquiries or questions from the teacher.						
69. Does not initiate conversations with other children.						
70. Hesitates a long time before making choices.						
71. Withdraws when teased by other children.						
	0	1	2	3	4	5
72. If not working well at the task assigned, drifts off and finds a way to comfort himself.						
73. Apologizes for himself/his behavior.						
74. Stutters.						
75. Utters non-sensical phrases or sentences.						
76. Comments that nobody likes him.						
77. Expresses worry or concern about bad grades, health, etc.						
78. Is absent from school when a major assignment or test is due.						
79. Drops an activity when he loses at that activity.						

	Rate of Occurrence					
	0	1	2	3	4	5
80. Appears tired and lethargic even though not suffering fatigue from physical activity.						
81. Distracted from the task at hand by ordinary classroom stimuli, minor noises, movements, etc.						
82. Remains in one position for long periods and stares fixedly while doing so.						
83. Loses interest in what he is doing and begins to disturb the class.						
84. Shows muscle irregularities, spasticity, rigidities.						
85. Does not take his turn in group activities.						
86. Comments that he is unhappy.						
87. Prefers to play with younger children even though children his own age are available.						
88. Comments that a particular activity is too hard for him and then quits.						
	0	1	2	3	4	5
89. States others are to blame for his actions.						
90. Does not pronounce words clearly.						
91. Tells stories which exaggerate the truth.						
92. Interrupts the class with comments which have no bearing on the class activity.						
93. Volunteers for classroom status assignments but does not finish them.						
94. Repeats same acts over and over in a mechanical fashion.						
95. When presented with a task, withdraws from the situation.						
96. Comments that he is stupid.						
97. Writes phrases in an immature fashion using large and badly formed letters.						
98. Complains of difficulty in breathing.						

Rate of Occurrence

	0	1	2	3	4	5
99. In structured physical activities, refuses to be a team leader if chosen for the position.						
100. Cries without apparent provocation.						
101. Requests praise or approval for tasks attempted.						
102. Comments that he does not feel well.						
103. Does not ask for directions to be repeated even when it is obvious he does not understand them.						
104. Is easily thrown off and makes errors.						
105. Mimics speech of others.						
106. Complains of others' unfairness toward him.						
107. Talks out of turn.						
108. Although he does not create a disturbance or disrupt the class, does not do any school work for given periods of time.						
	0	1	2	3	4	5
109. Shifts from one activity to the next without accomplishing either.						
110. Is hyperactive; e.g., constantly moving.						
111. Comments that he is tired.						
112. Gives excuses for not getting work in on time.						
113. Stumbles or falls.						
114. Cries whenever the teacher directs attention toward him.						
115. Must have things in perfect order.						
116. Reports difficulty in thinking; e.g., I can't concentrate.						
117. Seeks approval from teacher for tasks attempted.						
118. Uses baby talk.						

		Rate of Occurrence					
		0	1	2	3	4	5
119.	Comments that he is unable to complete a required classroom activity.						
120.	Talks to himself.						
121.	Answers questions about himself with "I don't know" or fails to answer.						
122.	Comments that others are out to get him or have it in for him.						
123.	Does not engage in group activities on the playground.						
124.	Displays poor coordination in physical activities.						

** Do not write on this page.

TALLY SHEET

I. Sub-Group Scores

Categories

		#1	#2	#3	#4	#5	#6	#7
Sections	A	_____	_____	_____	_____	_____	_____	_____
	B	_____	_____	_____				
	C	_____	_____	_____				
Subtotal		______	______	______	______	______	______	______

	#8
Section A	_____
Subtotal	_______

Grand Total ________

BEHAVIOR CLASSIFICATION SYSTEM

I. Scoring guide for behavioral categories contained in the Behavior Rating Scale.

Social Manifestations

A. Categories

1. Oppositional Behavior: Behaviors in this category are characterized by aggression expressed in oppositional behavior patterns of a generally passive character. These behaviors have a provocative quality associated with them which is expressed in the form of negativism, stubbornness, dawdling, procrastination, resistance, and defiance

Scale Items

1 ____
3. ____
5 ____
7. ____
9 ____
11. ____
13. ____
15 ____
17. ____
19 ____
21. ____
23 ____
25. ____
27 ____
29. ____
31. ____
33. ____
35. ____
37. ____
39. ____

____ % positive

2. Overt, aggressive behavior (verbal and physical) This category is defined by behaviors which generally involve an expenditure of energy. These behaviors represent overt, acting-out samples of behavior in which aggression is either goal directed (i.e. displays physical aggression toward persons or objects) or is a response to a specific environmental event

Scale Items

2. ____
4 ____
6 ____
8 ____
10 ____
12. ____
14 ____
16 ____
18 ____
20 ____
22. ____
24 ____
26 ____
28. ____
30. ____
32. ____
34. ____
36. ____
38. ____
40. ____
41. ____
43. ____
61. ____
62 ____
63. ____
64 ____

____ % positive

3. Deviations in social development. Behaviors in this category define behavior patterns in which an S experiences disturbed relationships with others (peers, parents, teachers, etc.) because of inappropriate responding to social stimuli.

Scale Items

42. ____
44. ____
45 ____
46 ____
47. ____
48 ____
49 ____
50. ____
51 ____
52 ____
53 ____
54. ____
55. ____
56. ____
57. ____
58. ____
59 ____
60. ____

____ % positive

Developmental Manifestations

4 Neurological/physical/motor manifestations Behaviors in this category are associated with the classic Strauss Syndrome and include the traditional symptoms of neurological impairment in addition to behaviors related to this syndrome such as minimal efficiency in learning, difficulty in writing, physical manifestations, etc.

Scale Items

66. ____
67. ____
80. ____
81. ____
82. ____
84. ____
97. ____
98. ____
102. ____
104. ____
110. ____
111. ____
113. ____
116. ____
124. ____

____ % positive

5. Signs of restricted functioning. This category describes children whose performance (physical, social, academic) is below the expectations of the school environment. Behaviors making up this category would be: confusion, daydreaming, extreme shyness, boredom, lack of flexibility in behavior.

Scale Items

68. ____
69 ____
70 ____
71. ____
72 ____
73 ____
83 ____
85 ____
87 ____
99. ____
100. ____
101. ____
103. ____
114 ____
115. ____
117. ____
121. ____
123 ____

____ % positive

6. Failure to follow through. Behaviors in this category are self-defining in that they refer to a general behavior pattern in which a number of tasks and activities are initiated by the subject but are seldom carried through to completion.

Scale Items

65. ____
78. ____
79. ____
93. ____
94. ____
95. ____
108. ____
109. ____
112. ____

____ % positive

Linguistic Manifestations

7. Verbal manifestations. Items in this category refer to behaviors which indicate an inadequate or inappropriate use of language (Immature defective speech or inappropriate verbal behavior.)

Scale Items

74. ____
75. ____
90. ____
91. ____
92. ____
105. ____
107. ____
118. ____
120. ____

____ % positive

8. Semantic negativism. This category is composed of behaviors which represent negative verbal statements that are usually self-directed, i.e., negative statements made about oneself.

Scale Items

76. ____
77. ____
86. ____
88. ____
89. ____
96. ____
106. ____
119. ____
122. ____

____ % positive

APPENDIX D

Name ______________________________ Date __________ Observer ________

Activity______________________________ Time __________ To ____________

TOI																													
TOD																													
NTD																													
H																													
D																													

TOI																													
TOD																													
NTD																													
H																													
D																													

DESCRIPTION OF BEHAVIORAL CATEGORIES

TOI = TASK-ORIENTED INDEPENDENT (Student completely involved in task)

TOD = TASK-ORIENTED DEPENDENT (Teacher assisted or waiting)

NTD = NON-TASK RELATED DEVIANCY (Behaviors disruptive of a learning climate - e.g., talking out, facial grimaces, etc.)

H = HAND (Seeking teacher assistance)

D = DISTRACTION (Non-task oriented; non-deviant - e.g., wandering about room, staring into space, sharpening pencils, going to lavatory, getting a drink, etc.)

ERIC REPORT RESUME

Resume Date 6 - 15 - 67	Is document copyrighted? ERIC Reproduction release?	Yes Yes

Title: Identification and Treatment of Social-Emotional Problems

Personal Authors: Mattson, Robert H., Walker, Hill M., Mattos, Robert L.

Institution (Source): University of Oregon, Eugene, Oregon, School of Education

Publication Date: 15 June 67 **Contract Number**: USOE Contract OEG 4-6-061308-0571

Pagination: 110 p.

Abstract:

The research project is divided into two sections: (1) The first section focuses upon developing assessment instruments for the identification of disturbed children. (2) The second section is concerned with developing a treatment model that will be effective in modifying the behavior of disturbed children in the educational setting.

A behavior checklist, a behavior rating scale, and a behavioral observation form have been constructed for the purpose of fulfilling objective one. Procedures on validating and estimating the reliability of the checklist have been completed. The split-half reliability estimate is .98. The scale discriminates between disturbed and non-disturbed children at the .001 level of confidence. Scores on the checklist ($N = 534$) correlate .68 with a criterion of behavior disturbance. Preliminary data on the rating scale indicates that the scale reflects treatment differences which are known to exist - $p = .01$. The average inter-rater reliability for three judges on the behavior of six subjects was .935. Agreement measures between independent observers using the behavioral observation form are .90 and above.

The treatment model, based upon learning theory principles, has produced measurable behavior change in disturbed fourth, fifth, and sixth grade male subjects. The researchers are not in a position at this time to indicate which treatment variables are producing a given amount of behavior change. The ensuing year will be spent in determining the weight which each specifiable treatment variable exerts upon the dependent variable of behavior change. Changes recorded to this writing indicates reduced frequencies of deviant behavior and increased proportions of time spent engaged in task oriented behavior.